PLASTID DIVISION AND RELATED GENES AND PROTEINS, AND METHODS OF USE

FIELD OF THE INVENTION

The present invention was funded in part with government support under grant number MCB 0092448 from the National Institutes of Health. The government may have certain rights in this invention.

The present invention relates to genes encoding proteins involved in prokaryotic-type or plastid division and/or morphology, and the encoded proteins, and in particular to isolated *Ftn2* (*ARC6*), *ARC5*, and *Fzo-like* genes and polypeptides. The present invention also provides methods for using *Ftn2* (*ARC6*), *ARC5*, and *Fzo-like* genes, and polypeptides.

BACKGROUND OF THE INVENTION

Plastids, the major organelles found only in plant and algal cells, are responsible for photosynthesis, for the storage of a wide variety of products, and for the synthesis of key molecules required for basic structural and functional aspects of plant cells. For example, plastids are responsible for the biosynthesis of purines and pyrimidines, and are the sole site of the synthesis of chlorophylls, carotenoids, certain amino acids (the "essential" amino acids), starches, fatty acids, and certain lipids.

Plastids are derived from proplastids, which are always present in young meristematic regions of a plant (a meristem is an undifferentiated region from which new cells arise). Proplastids can give rise to several different types of plastids, which types include: amyloplasts, unpigmented plastids which contain starch granules and which are especially common in storage organs, such as potato tubers; leucoplasts, colorless plastids involved in the synthesis of monoterpenes, the volatile compounds contained in essential oils and many of which are of commercial importance; chloroplasts, the green photosynthetic plastids responsible for energy capture via photosynthesis; and chromoplasts, yellow, orange, or red

plastids, depending upon the particular combination of carotenes and xanthopylls present, and which are responsible for the colors of many fruits (tomatoes, oranges), flowers (buttercups, marigolds) and roots (carrots, sweet potatoes).

Plastids arise from the binary fission of existing plastids, independently of cell division. In root tips, shoots, and other meristems, proplastid division keeps pace with cell division, so the daughter cells possess approximately the same number of plastids as the parent cells; in angiosperms, this number is about 20 proplastids per cell. As cell expansion supersedes cell division, the number of plastids per cell increases due to continued plastid division. The number of plastids present in a mature plant cell is typically similar for a particular cell in a particular tissue; for example, an Arabidopsis leaf mesophyll cell typically contains about 120 chloroplasts. Thus, plastid division is essential for the maintenance of plastid populations in plant cells undergoing division, and for the accumulation of large chloroplast numbers in photosynthetic tissues.

Plastids are surrounded by a double membrane system, which is made up of the outer and inner envelopes. The soluble interior portion of the plastid inside the inner envelope is the stroma; additional membrane structures may be present within the stroma, such as thylakoids. Thylakoids appear as interconnected stacked grana present in green chloroplasts, and contain the pigments necessary for light capture, such as chlorophyll. Thus, plastid division involves division of the outer and inner envelopes, as well as of the stroma and interior structures. As determined by ultra structural studies, plastid division begins with a constriction in the center of the plastid. Formation of the constriction is frequently associated with the appearance of an electron-dense annular structure termed the plastid dividing (PD) ring. In some electron micrographs of plastids from plants, the PD ring can be resolved into two concentric rings, an inner PD ring associated with the stromal surface of the inner envelope membrane, and an outer PD ring associated with the cytosolic surface of the outer envelope membrane. In other electron micrographs of plastids from red algae, yet a third PD ring is observed in the intermembrane space between the inner and outer envelope membranes. The constriction deepens and tightens, creating an extremely narrow isthmus before the two daughter plastids separate completely.

The mechanisms mediating plastid division are poorly understood, although it is believed that the PD rings are a dynamic macromolecular complex. It is also believed that this macromolecular complex is composed of numerous proteins that coordinate the mechanical activity required to constrict the plastid. Only a few components of the plastid division complex have been identified to date.

Plastid division is believed to have its evolutionary origin in a cyanobacterial endosymbiont that gave rise to chloroplasts (Osteryoung, KW et al. (1998) Plant Cell 10: 1991-2004). Thus, it has been proposed that the plastid division apparatus might have components in common with those involved in prokaryotic cell division, and in particular with cyanobacterial cell division (Possingham, JV and Lawrence ME (1983) Int. Rev. Cytol. 84: 1-56; and Suzuki, K et al (1994) J Cell Biol 63: 280-288). Genes from non-photosynthetic bacteria which play a role in division have been sequenced and identified. However, only a few of these genes involved in cyanobacterial division have been identified to date. One identified gene encodes bacterial FtsZ (from filamentation temperature-sensitive mutants, or fts mutants), which is a structural homologue to, and very likely the evolutionary precursor of, the eukaryotic tubulins (Erickson, HP (1998) Trends Cell Biol 7: 362-367; Faguy, DM and Doolittle WR (1998) Curr Biol 8: R338-341; Lowe, J and Amos LA (1998) Nature 391: 203-206) and Nogales, E et al. (1998) Nat Struct Biol 5: 451-458). FtsZ is well known to be a self-polymerizing, filament-forming GTPase, and it functions during bacterial cell division by assembling into a ring structure at the division site on the interior surface of the cytoplasmic membrane (Bi, E and Lutkenhaus J (1991) Nature 354: 161-164). The FtsZ ring assembly is required for the subsequent midcell localization of all other components of the cell division apparatus (Addinall, SG et al (1996) J Bacteriol 178: 3877-3884; and deBoer, PAJ et al. (1988) J Bacteriol 170: 2106-2112); it remains associated with the leading edge of the division septum throughout cytokinesis, then it disassembles immediately following cell separation before rapidly reassembling at the center of the newly formed daughter cells (Addinall, SG et al (1996) J Bacteriol 178: 3877-3884; Bi, E and Lutkenhaus J (1991) Nature 354: 161-164; Butterfass, T (1988) in Division and Segregation of Organelles (Cambridge, UK; Cambridge University Press) pp 21-38; and Sun, Q and Margolin, W (1998) J Bacteriol 180: 2020-2056). In *E. coli*, placement of the FtsZ ring is governed by the minB operon,

which encodes three gene products: MinC, MinD, and MinE (Lutkenhaus, J (1998) Curr Opin Microbiol 1: 210-215; Rothfield, L (1999) Annu Fev Genet 33: 423-448; Rothfield, LI and Justice, SS (1997) Cell 88: 581-584; and Sullivan, SM and Maddock, JR (2000) Curr Biol 10: R249-252).

FtsZ genes have also been found in nuclear genomes of land plants, as determined from plant gene database analysis. The encoded proteins fall into two major groups, FtsZ1 and FtsZ2 (Osteryoung KW, Stokes KD, Rutherford SM, Percival AL, and Lee, WY (1998), Plant Cell 10:1991-2004). FtsZ1 family proteins appear to contain cleavable chloroplast transit peptides at their amino terminal ends that target them to the chloroplast stromal compartment (Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) J. Mol. Biol. 300:1005-16), whereas members of the FtsZ2 family do not appear to possess easily recognizable chloroplast transit sequences. However, experimental evidence shows that both FtsZ1 and FtsZ2 proteins are imported into chloroplasts and localized in the stroma (McAndrew et al. (2001) *Plant Physiol.* 127:1656-1666). The FtsZ1 and FtsZ2 proteins are reported to colocalize to rings at the plastid midpoint in Arabidopsis and other plants, where members of both families assemble into rings on stromal surface of the inner envelope membranes (Osteryoung, KW and McAndrew, RS (2001) Annu Rev Plant Physiol Plant Mol Biol 52:315-333; and McAndrew et al. (2001) *Plant Physiol.* 127:1656-1666). These FtsZ proteins have been characterized both biochemically and microscopically during non-photosynthetic bacterial division; efforts are under way to similarly characterize these proteins in plants. (for a review, see Osteryoung, K and McAndrew RS (2002) Annu Rev Plant Physiol Mol Biol 52: 315-322; and McAndrew et al. (2001) *Plant Physiol.* 127:1656-1666). A MinD protein has also been found encoded in plastid genomes of algae, as well as in the nuclear genomes of higher plants (Colletti KS, Tatersall EA, Pyke KA, Froelich AE, Stokes KD, Osteryoung KW (2000) Curr. Biol. 10:507-16,Moehs CP, Tian L, Osteryoung KW, DelaPenna D (2001) Plant Mol. Biol. In press); at least some of the MinD proteins include a cleavable chloroplast target sequence (Osteryoung, K and McAndrew RS (2002) Annu Rev Plant Physiol Mol Biol 52: 315-322). Reduced expression of MinD in Arabidopsis plants results in plants with asymmetrically constricted plastids (Colletti KS, Tatersall EA, Pyke KA, Froelich AE, Stokes KD, Osteryoung KW (2000) Curr. Biol. 10:507-16), suggesting that

MinD also functions in plants to control the placement of the division ring to the center of the plastid. Both MinD as well as MinE are also encoded in the plastid genomes of unicellular algae (Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, et al. (1997) Proc. Natl. Acad. Sci. USA 94:5967-72).

Currently, *FtsZ*, *MinD*, and *MinE* are the only obvious homologues of non-photosynthetic bacterial cell division genes known to exist in photosynthetic eukaryotes, and roles for MinE and MinD in plastid division have only recently been demonstrated, where they are involved in placement of the PD rings at the site of plastid constriction (Itoh et al. (2001) *Plant Physiol.* 127:1644-1655; Reddy et al. (2002) *Planta.* 215:167-176). Even the function of most of the other non-photosynthetic bacterial cell division proteins are not well understood, and they therefore cannot provide clues as to whether functional counterparts participate in plastid division. However, at least nine proteins localize to the division septum in E. coli (Margolin W (1198)Trends Microbiol. 6:233-38, Rothfield LI, Justice SS (1997) Cell 88:581-84), and the plastid division apparatus is likely to be at least as complex (Osteryoung KW, Pyke KA (1998) Curr Opin. Plant Biol. 1:475-79).

Therefore, there is a need to identify and characterize other genes involved in plastid division. The discovery of such genes is useful to further characterize the mechanism of plastid division. Moreover, these genes can then be manipulated to vary the number and size of plastids present in plant cells, in order to vary agronomic and horticultural characteristics of economically important plants, such as crop, ornamental, and woody plants.

SUMMARY OF THE INVENTION

The present invention relates to compositions comprising Ftn2, ARC5, and Fzo-like genes and polypeptides. The present invention is not limited to any particular nucleic acid or amino acid sequence. The present invention also provides methods for using Ftn2, ARC5, and Fzo-like genes and polypeptides.

Thus, the present invention provides an isolated nucleic acid sequence comprising an Ftn2 gene. The present invention also provides an isolated nucleic acid sequence comprising a sequence encoding an Ftn2 polypeptide. In some embodiments, the Ftn2 gene product functions in division of a photosynthetic prokaryote or a plastid. In particular embodiments,

the nucleic acid sequence comprises SEQ ID NOs: 1, 3 or 4, or the coding sequence of SEQ ID NO:2.

The present invention also provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence comprising an Ftn2 gene. The present invention also provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence encoding an Ftn2 polypeptide. In some embodiments, a product of the first nucleic acid sequence functions in division of a photosynthetic prokaryote or a plastid. In particular embodiments, the second nucleic acid sequence is SEQ ID NOs: 1 or 4 or the coding sequence of SEQ ID NO:3.

The present invention also provides an isolated nucleic acid sequence comprising an Ftn2 gene, wherein the Ftn2 gene comprises at least one mutation. In some embodiments, the mutation is at least one nucleic acid substitution, nucleic acid addition, and/or nucleic acid deletion, and/or any combination of at least one nucleic acid substitution, nucleic acid addition, and/or nucleic acid deletion. The present invention also provides a nucleic acid sequence comprising an Ftn2 gene, where the gene encodes a variant of an Ftn2 polypeptide. In some embodiments, the variant is a mutant polypeptide, a truncated polypeptide, a fusion polypeptide, and/or any combination of a mutant polypeptide, a truncated polypeptide, and/or a fusion polypeptide. In particular embodiments, the isolated nucleic acid sequence comprises SEQ ID NO: 9 or the coding sequence of SEQ ID NO:10.

The present invention also provides an isolated antisense sequence corresponding to a nucleic acid sequence comprising an Ftn2 gene. The present invention also provides an isolated antisense sequence corresponding to a nucleic acid sequence encoding an Ftn2 polypeptide.

The present invention also provides an siRNA targeted to an RNA transcribed from an Ftn2 gene. The present invention also provides an siRNA targeted to an RNA transcribed from a nucleic acid sequence encoding an Ftn2 protein. The present invention also provides an isolated nucleic acid sequence encoding an siRNA targeted to an RNA transcribed from an Ftn2 gene. The present invention also provides an isolated nucleic acid sequence encoding an

siRNA targeted to an RNA transcribed from a nucleic acid sequence encoding an Ftn2 protein.

The present invention also provides compositions comprising any of the isolated nucleic acid sequences described above.

The present invention also provides any of the nucleic acid sequences described above operably linked to a heterologous promoter. The present invention also provides a vector comprising any of the nucleic acid sequences described above. In some embodiments, the vector comprises any of the nucleic acid sequences described above operably linked to a heterologous promoter.

The present invention also provides a purified protein, comprising an Ftn2 polypeptide. In some embodiments, the Ftn2 polypeptide functions in division of a photosynthetic prokaryote or a plastid. In particular embodiments, the protein comprises amino acid sequence SEQ ID NOs:2 or 4. The present invention also provides a purified protein, comprising a variant of an Ftn2 polypeptide. In some embodiments, the variant is a mutant polypeptide, a truncated polypeptide, a fusion polypeptide, and/or any combination of a mutant polypeptide, a truncated polypeptide, and/or a fusion polypeptide. In particular embodiments, the protein comprises amino acid sequence SEQ ID NO:11.

The present invention also provides compositions comprising any of the purified proteins described above.

The present invention also provides an organism transformed with any of the nucleic acid sequences described above. In some embodiments, the organism is a plant or a microorganism. In other embodiments, the present invention provides a plant transformed with any of the nucleic acid sequences described above. In yet other embodiments, the present invention provides a plant cell transformed with any of the nucleic acid sequences described above. In yet other embodiments, the present invention provides a plant seed transformed with any of the nucleic acid sequences described above. In particular embodiments, the nucleic acid sequence comprises SEQ ID NOs: 1 or 4 or the coding sequence of SEQ ID NO:3.

The present invention also provides an organism transformed with a heterologous gene comprising an Ftn2 gene. In some embodiments, the organism is a plant or a microorganism.

In other embodiments, the present invention provides a plant transformed with a heterologous gene comprising an Ftn2 gene. In yet other embodiments, the present invention provides a plant cell transformed with a heterologous gene comprising an Ftn2 gene. In yet other embodiments, the present invention provides a plant seed transformed with a heterologous gene comprising an Ftn2 gene. In particular embodiments, the nucleic acid sequence comprises SEQ ID NOs: 1 or 4 or the coding sequence of SEQ ID NO:3.

In additional embodiments, the present invention provides an isolated nucleic acid sequence comprising an ARC5 gene. In some embodiments, the present invention provides an isolated nucleic acid sequence comprising a sequence encoding an ARC5 polypeptide. In some embodiments, the ARC5 gene is selected from the group consisting of SEQ ID NOs: 11 and 14. In some embodiments, ARC5 polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 13, 16, 17, and 18. In other embodiments, the present invention provides an isolated antisense sequence corresponding to a nucleic acid sequence comprising an ARC5 gene. In still other embodiments, the present invention provides an isolated antisense sequence corresponding to a nucleic acid sequence encoding an ARC5 polypeptide. In still further embodiments, the present invention provides an siRNA targeted to an RNA transcribed from an ARC5 gene. In yet other embodiments, the present invention provides an siRNA targeted to an RNA transcribed from a nucleic acid sequence encoding an ARC5 protein.

The present invention also provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence comprising an ARC5 gene. In some embodiments, a product of the first nucleic acid sequence functions in division of a photosynthetic prokaryote or a plastid.

The present invention additionally provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence encoding an ARC5 polypeptide. In some embodiments, a product of the first nucleic acid sequence functions in division of a photosynthetic prokaryote or a plastid. In some embodiments, the second nucleic acid sequence is SEQ ID NO: 11 or 14.

In still further embodiments, the present invention provides an isolated nucleic sequence comprising an ARC5 gene, wherein the ARC5 gene comprises at least one mutation.

In some embodiments, the mutation is at least one nucleic acid substitution, addition, deletion, and/or any combination of at least one nucleic acid substitution, addition, and/or deletion.

In certain embodiments, the present invention provides a ARC5 nucleic acid sequence operably linked to a heterologous promoter. In some embodiments, the present invention provides vector comprising an ARC5 nucleic acid sequence. In other embodiments, the present invention provides a vector comprising an ARC5 nucleic acid sequence operably linked to a heterologous promoter.

In some embodiments, the present invention provides an isolated protein, comprising an ARC5 polypeptide; in particular embodiments, the ARC5 polypeptide comprises amino acid sequence SEQ ID NO:13, 16, 17, or 18. In other embodiments, the present invention provides an isolated protein, comprising a variant of an ARC5 polypeptide. In some embodiments, the variant is a mutant polypeptide, a truncated polypeptide, a fusion polypeptide, and/or any combination of a mutant polypeptide, a truncated polypeptide, and/or a fusion polypeptide.

In certain embodiments, the present invention provides an organism transformed with a heterologous gene comprising an ARC5 gene. In some embodiments, the organism includes, but is not limited to, a plant, an algae, or a microorganism. In other embodiments, the present invention provides a plant, a plant cell, or a plant seed transformed with a heterologous gene comprising an ARC5 gene. The present invention also provides an organism transformed with a heterologous gene encoding an ARC5 polypeptide, and a plant, plant cell, or plant seed transformed with a heterologous gene encoding an ARC5 polypeptide.

In additional embodiments, the present invention provides an isolated nucleic acid sequence comprising an Fzo-like gene. In some embodiments, the present invention provides an isolated nucleic acid sequence comprising a sequence encoding an Fzo-like polypeptide. In some embodiments, the Fzo-like gene is selected from the group consisting of SEQ ID NOs: 19 and 22. In some embodiments, the Fzo-like gene further comprises the nucleic acid sequence of SEQ ID NO:25 at the 3' terminus. In some embodiments, Fzo-like polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 21 or 24. In other embodiments, the present invention provides an isolated antisense sequence corresponding to a nucleic acid sequence comprising an Fzo-like gene. In still other

embodiments, the present invention provides an isolated antisense sequence corresponding to a nucleic acid sequence encoding an Fzo-like polypeptide. In still further embodiments, the present invention provides an siRNA targeted to an RNA transcribed from an Fzo-like gene. In yet other embodiments, the present invention provides an siRNA targeted to an RNA transcribed from a nucleic acid sequence encoding an Fzo-like protein.

The present invention also provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence comprising an Fzo-like gene. In some embodiments, a product of the first nucleic acid sequence functions in division of a photosynthetic prokaryote or a plastid.

The present invention additionally provides an isolated first nucleic acid sequence that hybridizes under conditions of high stringency to a second nucleic acid sequence encoding an Fzo-like polypeptide. In some embodiments, a product of the first nucleic acid sequence functions in division of a photosynthetic prokaryote or a plastid. In some embodiments, the second nucleic acid sequence is SEQ ID NO: 19 or 22. In some embodiments, the Fzo-like nucleic acid further comprises the nucleic acid sequence of SEQ ID NO:25 at the 3' terminus.

In still further embodiments, the present invention provides an isolated nucleic sequence comprising an Fzo-like gene, wherein the Fzo-like gene comprises at least one mutation. In some embodiments, the mutation is at least one nucleic acid substitution, addition, deletion, and/or any combination of at least one nucleic acid substitution, addition, and/or deletion.

In certain embodiments, the present invention provides a Fzo-like nucleic acid sequence operably linked to a heterologous promoter. In some embodiments, the present invention provides vector comprising an Fzo-like nucleic acid sequence. In other embodiments, the present invention provides a vector comprising an Fzo-like nucleic acid sequence operably linked to a heterologous promoter.

In some embodiments, the present invention provides an isolated protein, comprising an Fzo-like polypeptide; in particular embodiments, the Fzo-like polypeptide comprises amino acid sequence SEQ ID NO:21 or 24. In other embodiments, the present invention provides an isolated protein, comprising a variant of an Fzo-like polypeptide. In some embodiments, the variant is a mutant polypeptide, a truncated polypeptide, a fusion

polypeptide, and/or any combination of a mutant polypeptide, a truncated polypeptide, and/or a fusion polypeptide.

In certain embodiments, the present invention provides an organism transformed with a heterologous gene comprising an Fzo-like gene. In some embodiments, the organism includes, but is not limited to, a plant, an algae, or a microorganism. In other embodiments, the present invention provides a plant, a plant cell, or a plant seed transformed with a heterologous gene comprising an Fzo-like gene. The present invention also provides an organism transformed with a heterologous gene encoding an Fzo-like polypeptide, and a plant, plant cell, or plant seed transformed with a heterologous gene encoding an Fzo-like polypeptide

DESCRIPTION OF THE FIGURES

Figure 1 shows nucleic acid sequences of AtFtn2 (ARC6 gene) from a wild type plant in a WS ecotype and of arc6-1 gene in an arc6-1 mutant plant in a WS-like ecotype. Panel A shows a cDNA sequence (SEQ ID NO:1), and panel B shows a genomic sequence (SEQ ID NO:3) of AtFtn2 gene; panel C shows a cDNA sequence (SEQ ID NO:9) and panel D shows a genomic sequence (SEQ ID NO:10) of arc6-1 gene.

Figure 2 shows the amino acid sequences of the peptide encoded by AtFtn2 (ARC6 gene) from a wild type plant in a WS ecotype (panel A, SEQ ID NO:2) and of the peptide encoded by arc6-1 gene in an arc6-1 mutant plant in a WS-like ecotype (panel B, SEQ ID NO:11).

Figure 3 shows the structure of the *AtFtn2* gene (Panel A) and protein (Panel B). Panel A shows that the open reading frame is terminated by a TAA in-frame stop codon. The diagram depicts introns (thin lines) and exons (black boxes). Sizes are given in bp. The position of the *arc6-1* mutation (C -> T) at position 1141 is marked. The nucleotide sequences flanking the mutation (underlined) show the change of codon 325 (CGA in a wild type plant) into a premature stop (TGA) in *arc6-1*. Panel B shows the putative functional and conserved protein domain, which are depicted as wider black boxes; their numerical positions within the AtFtn2 sequence are also indicated. Black lines above the diagram delineate regions of AtFtn2 conserved among Ftn2 homologues (see Figures 4-6). CT, chloroplast targeting signal.

Figure 4 shows a sequence alignment of DnaJ-like domains of plant and cyanobacterial Ftn2 proteins (indicated by asterisk) and DnaJ domains from Pfam database. Total about 270 DnaJ domains from the database were aligned with the ARC6 proteins. Shown in this figure are only selected DnaJ domains most similar to Ftn2 proteins. Black and gray columns indicate that identical or similar amino acid, respectively, was present in 70% of all aligned sequences at that position. The TrEMBL accession codes and location of the DnaJ domain within the protein are shown for the Pfam database records. For the ARC6 homologues, if the protein sequences were derived from EST records and did not encompass the initial M, the location of the DnaJ domain is not given.

Figure 5 shows an alignment of plant and cyanobacterial Ftn2 full and partial sequences. Partial sequences are marked by asterisk (*). Not shown are the N-termini of the plant sequences, which contain chloroplast transit peptides. Light-gray and black columns indicate similarity and identity, respectively, greater than 80%. Gaps are indicated by a dash (-), missing sequence by an underline (_). Similarity and identity calculations do not include missing sequences. The Dna-J like domain is indicated by a solid line ( ) Putative myb domain is indicated by diamonds ( ). Site of truncation of the protein in *arc6* mutant is marked by a triangle ( ) at position 398 of the alignment (residue 325 of AtFtn2).

Figure 6 shows the nucleotide sequence (panel A , SEQ ID NO:4) and amino acid sequence (panel B , SEQ ID NO:5) of *ftn2* from *Synechococcus* sp. PCC 7942; these sequences have been submitted to GenBank under accession no. AF21196.

Figure 7 shows the nucleotide sequence (panel A , SEQ ID NO:6) and amino acid sequence (panel B, SEQ ID NO:7) of *ftn6* from *Synechococcus* sp. PCC 7942; these sequences have been submitted to GenBank under accession no. AF21197.

Figure 8 shows nucleotide and amino acid sequences of Ftn2 homologs described in Table 3.

Figure 9 shows the nucleic acid sequence of SEQ ID NO:11.

Figure 10 shows the nucleic acid sequence of SEQ ID NO:12.

Figure 11 shows the amino acid sequence of SEQ ID NO:13.

Figure 12 shows the nucleic acid sequence of SEQ ID NO:14.

Figure 13 shows the nucleic acid sequence of SEQ ID NO:15.

Figure 14 shows the amino acid sequence of SEQ ID NO:16.

Figure 15 shows the amino acid sequence of SEQ ID NO:17.

Figure 16 shows the amino acid sequence of SEQ ID NO:18.

Figure 17 shows the nucleic acid sequence of SEQ ID NO:19.

Figure 18 shows the nucleic acid sequence of SEQ ID NO:20.

Figure 19 shows the amino acid sequence of SEQ ID NO:21.

Figure 20 shows the nucleic acid sequence of SEQ ID NO:22.

Figure 21 shows the nucleic acid sequence of SEQ ID NO:23.

Figure 22 shows the amino acid sequence of SEQ ID NO:24.

Figure 23 shows the nucleic acid sequence of SEQ ID NO:25.

Figure 24 shows the genomic sequence of AtFzo-like gene. The sequences is the reverse complementary sequence; stop and start codons are indicated by underlined bold text SEQ ID NO:26 is the genomic sequence; SEQ ID NO:27 comprises the sequence between and including the stop and start codons.

Figure 25 shows an alignment of the AtARC5 gene with Dynamin-1 from *Homo sapiens* and Dnm1p from *Saccharomyces cerevisiae*. Gray boxes indicate completely conserved residues; yellow boxes are identical residues; cyan boxes are similar residues; dashes indicate gaps. The domain structure is indicated by the lines above the alignment. Red, GTPase domain; green, middle domain; blue, PH domain; lavender, GTPase effector domain; black, PR domain. The dotted underline indicates the sequence encoded by the alternatively spliced intron in *ARC5*. The triangle indicates the position of the *arc5* mutation.

Figure 26 shows additional sequences which are homologous to *AtARC5* gene.

Figure 27 shows additional sequences which are homologous to *AtFzo-like* gene.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases as used herein are defined below:

The term "plant" is used in it broadest sense. It includes, but is not limited to, any species of woody, ornamental or decorative, crop or cereal, fruit or vegetable plant, and photosynthetic green algae (*e.g., Chlamydomonas reinhardtii*). It also refers to a plurality of

plant cells that are largely differentiated into a structure that is present at any stage of a plant's development. Such structures include, but are not limited to, a fruit, shoot, stem, leaf, flower petal, etc. The term "plant tissue" includes differentiated and undifferentiated tissues of plants including those present in roots, shoots, leaves, pollen, seeds and tumors, as well as cells in culture (*e.g.*, single cells, protoplasts, embryos, callus, etc.). Plant tissue may be in planta, in organ culture, tissue culture, or cell culture. The term "plant part" as used herein refers to a plant structure or a plant tissue.

The term "crop" or "crop plant" is used in its broadest sense. The term includes, but is not limited to, any species of plant or algae edible by humans or used as a feed for animals or used, or consumed by humans, or any plant or algae used in industry or commerce.

The term "oil-producing species" refers to plant species which produce and store triacylglycerol in specific organs, primarily in seeds. Such species include but are not limited to soybean (*Glycine max*), rapeseed and canola (including *Brassica napus* and *B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*), coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*) and peanut (*Arachis hypogaea*). The group also includes non-agronomic species which are useful in developing appropriate expression vectors such as tobacco, rapid cycling *Brassica* species, and *Arabidopsis thaliana*, and wild species.

The term plant cell "compartments" or "organelles" is used in its broadest sense. The term includes but is not limited to, the endoplasmic reticulum, Golgi apparatus, trans Golgi network, plastids, sarcoplasmic reticulum, glyoxysomes, mitochondrial, chloroplast, and nuclear membranes, and the like.

The term "host cell" refers to any cell capable of replicating and/or transcribing and/or translating a heterologous gene.

The term "*arc*" refers to mutations observed in Arabidopsis which exhibition abnormal chloroplast accumulation and/or replication, and is an abbreviation for the designation " accumulation and replication of chloroplasts." Different arc mutants have been observed, and are indicated by a number after the *arc* designation: for example, *arc1*, *arc2*, etc.

The term "*Ftn2*" refers to a gene that when naturally occurring in a wild-type organism encodes an Ftn2 polypeptide. An Ftn2 polypeptide functions in prokaryotic-type division, such that a decreased amount of Ftn2 polypeptide in a prokaryote or a plant or algal cell compared to the amount typically present in wild-type results in incomplete division or no division of the prokaryote or plastid(s) in the plant or algal cell. As an illustrative but non-limiting example, in photosynthetic prokaryotes such as cyanobacteria, a decreased amount of Ftn2 polypeptide can result in long filamentous cells, up to many times longer than a wild-type cell. As an illustrative but non-limiting example, in plants such as Arabidopsis, a decreased amount of Ftn2 polypeptide can result in a single or a few very large chloroplasts present in a single leaf mesophyll cell.

An Ftn2 polypeptide is a protein (about 660 to about 800 amino acids long) which can be roughly defined by three regions. The N-terminal (about 420 amino acids) contains the DnaJ-like domain, and exhibits a high degree of homology among Ftn2 proteins obtained from different sources (about 20 to about 60% identity, and about 50 to about 80% similarity). The large central region (about 200 amino acids) is fairly variable, and exhibits a lower degree of homology among the different Ftn2 proteins (about 6% to about 20% identity, and about 20 to about 44% similarity). The C-terminal region (about 110 amino acids) is more highly conserved and in Arabidopsis Ftn2, contains putative myb domain (residues 677-690). The C-terminal region exhibits a higher degree of homology than the central region (about 15% to about 55% identity, and about 40 to about 70% similarity). The result is that when considered as a whole, homologous Ftn2 proteins possess about 15% or greater identity and about 38% or greater similarity to AtFtn2 protein. However, the N-terminal and C-terminal regions possess a higher degree of similarity and a higher degree of identity among the different Ftn2 proteins than do the whole proteins.

In Arabidopsis, a mutation in the Ftn2 gene results in an *arc* (accumulation and replication of chloroplasts) mutant, the *arc6* mutant. The evidence described in Example 2, including the observations that the sequences of Ftn2 from a wild-type background and the sequences of *arc6-1*, *arc6-2*, and *arc6-3* are essentially the same except that the a C -> T transition at position 1141 in the gene results in a premature stop codon and results in a

truncated protein of about 324 amino acids, and that the *arc6* mutant is rescued by a wild-type copy of AtFtn2, indicates that *AtFtn2* gene is *ARC6*.

The term "*ARC5*" refers to a gene that when naturally occurring in a wild-type organism encodes an ARC5 polypeptide. An ARC5 polypeptide functions in prokaryotic-type division, such that a decreased amount of ARC5 polypeptide in a prokaryote or a plant (including an algal) cell compared to the amount typically present in wild-type results in incomplete division or no division of the prokaryote or plastid(s) in the plant (including an algal) cell. As an illustrative but non-limiting example, in plants such as Arabidopsis, a decreased amount of ARC5 polypeptide can result in cells with about 5 to 10 chloroplasts per cell, where the chloroplasts are larger than in wild type, and constricted chloroplasts were frequently found.

An ARC5 polypeptide is a protein (of about 777 or about 741 amino acids long) which can be roughly defined by three regions. These regions, or motifs, are also found in other dynamin-like proteins: a conserved N-terminal GTPase domain, a pleckstrin homology (PH) domain shown in some proteins to mediate membrane association, and a C-terminal GTPase Effector Domain (GED) thought to interact directly with the GTPase domain and to mediate self-assembly.

In Arabidopsis, a mutation in the ARC5 gene results in an *arc* (accumulation and replication of chloroplasts) mutant, the *arc5* mutant, as described in Example 6. Moreover, in Arabidopsis, two distinct cDNAs encoding ARC5 proteins with uninterrupted reading frames of 777 (87.2 kDa) or 741 (83.5 kDa) amino acids are found. These results indicate that the *ARC5* transcript is alternatively spliced.

The term "*Fzo-like*" refers to a gene that when naturally occurring in a wild-type organism encodes an Fzo-like polypeptide. An Fzo-like polypeptide functions in prokaryotic-type division and/or morphology, such that a decreased amount of an Fzo-like polypeptide in a prokaryote or a plant (including an algal) cell compared to the amount typically present in wild-type results in incomplete division or no division and/or an abnormal morphology of the prokaryote or plastid(s) in the plant (including an algal) cell. As an illustrative but non-limiting example, in plants such as Arabidopsis, a T-DNA insertion in an *Fzo-like* gene can result in abnormalities in chloroplast size and number. Fzo-like polypeptide amino acid

sequences are similar to the yeast Fzo1, which functions in the control of mitochondrial morphology in yeast. Fzo-like polypeptides are contemplated to comprise several domains: a chloroplast transit peptide, a GTPase domain and two predicted trans-membrane domains. In Arabidopsis Fzo-like polypeptide, the predicted chloroplast transit peptide is the first 54 amino acids, the GTPase domain is between amino acids 350-500, and the two predicted trans-membrane domains are close to each other in the region between amino acids 770-830.

It is contemplated that *Ftn2, ARC5, and Fzo-like* genes and proteins are present in, and thus can be isolated from and/or used in, any organism which possesses plastids, as well as any photosynthetic bacteria such as cyanobacteria; organisms which posses plastids include plants, both vascular and non-vascular, algae, and some parasitic protists which contain vestigial plastids.

The term "prokaryotic-type division" refers to division of a prokaryote, and in particular of a photosynthetic prokaryote, or of a plastid.

The term "morphology" refers to the form and/or structure of an organism, an organ, a tissue, a cell, an organelle, or a subcellular structure (for example, a membrane), and its development, and in particular to the form and/or structure and development of the form and/or structure of plastids in plants.

The terms "protein" and "polypeptide" refer to compounds comprising amino acids joined via peptide bonds and are used interchangeably.

As used herein, where "amino acid sequence" is recited herein to refer to an amino acid sequence of a protein molecule, "amino acid sequence" and like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule; furthermore, an "amino acid sequence" can be deduced from the nucleic acid sequence encoding the protein.

The term "portion" when used in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino sequence minus one amino acid.

The term "homology" when used in relation to amino acids refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). "Sequence identity" refers to a measure of relatedness between two or more proteins, and is

given as a percentage with reference to the total comparison length. The identity calculation takes into account those amino acid residues that are identical and in the same relative positions in their respective larger sequences. Calculations of identity may be performed by algorithms contained within computer programs.

The term "chimera" when used in reference to a polypeptide refers to the expression product of two or more coding sequences obtained from different genes, that have been cloned together and that, after translation, act as a single polypeptide sequence. Chimeric polypeptides are also referred to as "hybrid" polypeptides. The coding sequences includes those obtained from the same or from different species of organisms.

The term "fusion" when used in reference to a polypeptide refers to a chimeric protein containing a protein of interest joined to an exogenous protein fragment (the fusion partner). The fusion partner may serve various functions, including enhancement of solubility of the polypeptide of interest, as well as providing an "affinity tag" to allow purification of the recombinant fusion polypeptide from a host cell or from a supernatant or from both. If desired, the fusion partner may be removed from the protein of interest after or during purification.

The term "homolog" or "homologous" when used in reference to a polypeptide refers to a high degree of sequence identity between two polypeptides, or to a high degree of similarity between the three-dimensional structure or to a high degree of similarity between the active site and the mechanism of action. In a preferred embodiment, a homolog has a greater than 60% sequence identity, and more preferable greater than 75% sequence identity, and still more preferably greater than 90% sequence identity, with a reference sequence.

The terms "variant" and "mutant" when used in reference to a polypeptide refer to an amino acid sequence that differs by one or more amino acids from another, usually related polypeptide. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (*e.g.*, replacement of leucine with isoleucine). More rarely, a variant may have "non-conservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Similar minor variations may also include amino acid deletions or insertions (i.e., additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological activity may be

found using computer programs well known in the art, for example, DNAStar software. Variants can be tested in functional assays. Preferred variants have less than 10%, and preferably less than 5%, and still more preferably less than 2% changes (whether substitutions, deletions, and so on).

The term "gene" refers to a nucleic acid (*e.g.*, DNA or RNA) sequence that comprises coding sequences necessary for the production of an RNA, or a polypeptide or its precursor (*e.g.*, proinsulin). A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, etc.) of the polypeptide are retained. The term "portion" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides to the entire gene sequence minus one nucleotide. Thus, "a nucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

The term "gene" also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences

are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

The term "heterologous gene" refers to a gene encoding a factor that is not in its natural environment (*i.e.*, has been altered by the hand of man). For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.*g., mutated, added in multiple copies, linked to a non-native promoter or enhancer sequence, etc.). Heterologous genes may comprise plant gene sequences that comprise cDNA forms of a plant gene; the cDNA sequences may be expressed in either a sense (to produce mRNA) or anti-sense orientation (to produce an anti-sense RNA transcript that is complementary to the mRNA transcript). Heterologous genes are distinguished from endogenous plant genes in that the heterologous gene sequences are typically joined to nucleotide sequences comprising regulatory elements such as promoters that are not found naturally associated with the gene for the protein encoded by the heterologous gene or with plant gene sequences in the chromosome, or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof.

The term "an oligonucleotide having a nucleotide sequence encoding a gene" or "a nucleic acid sequence encoding" a specified polypeptide refers to a nucleic acid sequence comprising the coding region of a gene or in other words the nucleic acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (*i.e.*,

the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, *etc.* may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, *etc.* or a combination of both endogenous and exogenous control elements.

The terms "complementary" and "complementarity" refer to polynucleotides (*i.e.,* a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). "Sequence identity" refers to a measure of relatedness between two or more nucleic acids, and is given as a percentage with reference to the total comparison length. The identity calculation takes into account those nucleotide residues that are identical and in the same relative positions in their respective larger sequences. Calculations of identity may be performed by algorithms contained within computer programs such as "GAP" (Genetics Computer Group, Madison, Wis.) and "ALIGN" (DNAStar, Madison, Wis.). A partially complementary sequence is one that at least partially inhibits (or competes with) a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for

and inhibit the binding (*i.e.*, the hybridization) of a sequence which is completely homologous to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

Low stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

It is well known that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency

hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.*g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low to high stringency as described above.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low to high stringency as described above.

The term "hybridization" refers to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

The term "T_m" refers to the "melting temperature" of a nucleic acid. The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (*See e.g.*, Anderson and Young, Quantitative Filter Hybridization (1985) *in Nucleic Acid Hybridization*). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein the term "stringency" refers to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid base

pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q βreplicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.* (1972) Proc. Natl. Acad. Sci. USA, 69:3038). Other nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin *et al.* (1970) Nature, 228:227). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace (1989) Genomics, 4:560). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.) (1989) *PCR Technology*, Stockton Press).

The term "amplifiable nucleic acid" refers to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

The term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

The term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

The term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase

extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

The terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

The term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

The term "reverse-transcriptase" or "RT-PCR" refers to a type of PCR where the starting material is mRNA. The starting mRNA is enzymatically converted to complementary DNA or "cDNA" using a reverse transcriptase enzyme. The cDNA is then used as a "template" for a "PCR" reaction.

The term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (*e.g.*, mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (*i.e.*, via the enzymatic action of an RNA polymerase), and into protein, through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products (*i.e.*, RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. Molecules (*e.g.*, transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and "repressors," respectively.

The terms "in operable combination", "in operable order" and "operably linked" refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

The term "regulatory element" refers to a genetic element which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, *etc.*

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis, *et al.*, Science 236:1237, 1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect, mammalian and plant cells. Promoter and enhancer elements have also been isolated from viruses and analogous control elements, such as promoters, are also found in prokaryotes. The selection of a particular promoter and enhancer depends on the cell type used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types (*for review, see* Voss, *et al.*, Trends Biochem. Sci., 11:287, 1986; and Maniatis, *et al.*, *supra* 1987).

The terms "promoter element," "promoter," or "promoter sequence" as used herein, refer to a DNA sequence that is located at the 5' end (*i.e.* precedes) the protein coding region of a DNA polymer. The location of most promoters known in nature precedes the transcribed region. The promoter functions as a switch, activating the expression of a gene. If the gene is activated, it is said to be transcribed, or participating in transcription. Transcription involves the synthesis of mRNA from the gene. The promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA.

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (*e.g.*, seeds) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (*e.g.*, leaves). Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant such that the reporter construct is integrated into every tissue of the resulting transgenic plant, and detecting the expression of the reporter gene (*e.g.*, detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a promoter refers to a promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, *e.g.*, immunohistochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the

promoter. A labeled (*e.g.*, peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (*e.g.*, with avidin/biotin) by microscopy.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. Exemplary constitutive plant promoters include, but are not limited to SD Cauliflower Mosaic Virus (CaMV SD; *see e.g.*, U.S. Pat. No. 5,352,605, incorporated herein by reference), mannopine synthase, octopine synthase (ocs), superpromoter (*see e.g.*, WO 95/14098), and *ubi3* (*see e.g.*, Garbarino and Belknap (1994) Plant Mol. Biol. 24:119-127) promoters. Such promoters have been used successfully to direct the expression of heterologous nucleic acid sequences in transformed plant tissue.

In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nuclei acid sequence in the presence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The enhancer and/or promoter may be "endogenous" or "exogenous" or "heterologous." An "endogenous" enhancer or promoter is one that is naturally linked with a given gene in the genome. An "exogenous" or "heterologous" enhancer or promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (*i.e.*, molecular biological techniques) such that transcription of the gene is directed by the linked enhancer or promoter. For example, an endogenous promoter in operable combination with a first gene can be isolated, removed, and placed in operable combination with a second gene, thereby making it a "heterologous promoter" in operable combination with the second gene. A variety of such combinations are contemplated (*e.g.*, the first and second genes can be from the same species, or from different species.

The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript in eukaryotic host cells. Splicing signals

mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook, *et al.* (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York, pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term "poly(A) site" or "poly(A) sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable, as transcripts lacking a poly(A) tail are unstable and are rapidly degraded. The poly(A) signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous poly(A) signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly(A) signal is one which has been isolated from one gene and positioned 3' to another gene. A commonly used heterologous poly(A) signal is the SV40 poly(A) signal. The SV40 poly(A) signal is contained on a 237 bp *Bam*HI/*Bcl*I restriction fragment and directs both termination and polyadenylation (Sambrook, *supra*, at 16.6-16.7).

The term "selectable marker" refers to a gene which encodes an enzyme having an activity that confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed, or which confers expression of a trait which can be detected (*e.g.*., luminescence or fluorescence). Selectable markers may be "positive" or "negative." Examples of positive selectable markers include the neomycin phosphotransferase (NPTII) gene which confers resistance to G418 and to kanamycin, and the bacterial hygromycin phosphotransferase gene (*hyg*), which confers resistance to the antibiotic hygromycin. Negative selectable markers encode an enzymatic activity whose expression is cytotoxic to the cell when grown in an appropriate selective medium. For example, the HSV-*tk* gene is commonly used as a negative selectable marker. Expression of the HSV-*tk* gene in cells grown in the presence of gancyclovir or acyclovir is cytotoxic; thus, growth of cells in

selective medium containing gancyclovir or acyclovir selects against cells capable of expressing a functional HSV TK enzyme.

The term "vector refers to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vehicle" is sometimes used interchangeably with "vector."

The terms "expression vector" or "expression cassette" refer to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The term "transfection" refers to the introduction of foreign DNA into cells. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, glass beads, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, viral infection, biolistics (*i.e.*, particle bombardment) and the like.

The terms "infecting" and "infection" when used with a bacterium refer to co-incubation of a target biological sample, (*e.g.*, cell, tissue, *etc.*) with the bacterium under conditions such that nucleic acid sequences contained within the bacterium are introduced into one or more cells of the target biological sample.

The term "*Agrobacterium*" refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term "*Agrobacterium*" includes, but is not limited to, the strains *Agrobacterium tumefaciens*, (which typically causes crown gall in infected plants), and *Agrobacterium rhizogens* (which causes hairy root disease in infected host plants). Infection of a plant cell with *Agrobacterium* generally results in the production of opines (*e.g.*, nopaline, agropine, octopine etc.) by the infected cell. Thus, *Agrobacterium* strains which cause production of nopaline (*e.g.*, strain LBA4301, C58, A208, GV3101) are referred to as "nopaline-type" *Agrobacteria*; *Agrobacterium* strains which cause production of octopine (*e.g.*, strain LBA4404, Ach5, B6) are referred to as "octopine-type"

Agrobacteria; and *Agrobacterium* strains which cause production of agropine (*e.*g., strain EHA105, EHA101, A281) are referred to as "agropine-type" *Agrobacteria*.

The terms "bombarding, "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (*e.g.*, cell, tissue, *etc.*) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (*e.g.*, U.S. Patent No. 5,584,807, the contents of which are incorporated herein by reference), and are commercially available (*e.g.*, the helium gas-driven microprojectile accelerator (PDS-1000/He, BioRad).

The term "microwounding" when made in reference to plant tissue refers to the introduction of microscopic wounds in that tissue. Microwounding may be achieved by, for example, particle bombardment as described herein.

The term "transgenic" when used in reference to a plant or fruit or seed (*i.e.*, a "transgenic plant" or "transgenic fruit" or a "transgenic seed") refers to a plant or fruit or seed that contains at least one heterologous gene in one or more of its cells. The term "transgenic plant material" refers broadly to a plant, a plant structure, a plant tissue, a plant seed or a plant cell that contains at least one heterologous gene in one or more of its cells.

The terms "transformants" or "transformed cells" include the primary transformed cell and cultures derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

The term "wild-type" when made in reference to a gene refers to a gene which has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product which has the characteristics of a gene product isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product which displays modifications in sequence and/or functional properties (*i.e.*, altered

characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "antisense" refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a sense strand of a DNA duplex. A "sense strand" of a DNA duplex refers to a strand in a DNA duplex which is transcribed by a cell in its natural state into a "sense mRNA." Thus an "antisense" sequence is a sequence having the same sequence as the non-coding strand in a DNA duplex. The term "antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene by interfering with the processing, transport and/or translation of its primary transcript or mRNA. The complementarity of an antisense RNA may be with any part of the specific gene transcript, *i.e.*, at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. In addition, as used herein, antisense RNA may contain regions of ribozyme sequences that increase the efficacy of antisense RNA to block gene expression. "Ribozyme" refers to a catalytic RNA and includes sequence-specific endoribonucleases. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of preventing the expression of the target protein.

The term "overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. The term "cosuppression" refers to the expression of a foreign gene which has substantial homology to an endogenous gene resulting in the suppression of expression of both the foreign and the endogenous gene. The term "altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

The term "recombinant" when made in reference to a nucleic acid molecule refers to a nucleic acid molecule which is comprised of segments of nucleic acid joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein molecule which is expressed using a recombinant nucleic acid molecule.

The terms "Southern blot analysis" and "Southern blot" and "Southern" refer to the analysis of DNA on agarose or acrylamide gels in which DNA is separated or fragmented according to size followed by transfer of the DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then exposed to a labeled probe to detect DNA species complementary to the probe used. The DNA may be cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during transfer to the solid support. Southern blots are a standard tool of molecular biologists (J. Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY, pp 9.31-9.58).

The term "Northern blot analysis" and "Northern blot" and "Northern" as used herein refer to the analysis of RNA by electrophoresis of RNA on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists (J. Sambrook, *et al.* (1989) *supra*, pp 7.39-7.52).

The terms "Western blot analysis" and "Western blot" and "Western" refers to the analysis of protein(s) (or polypeptides) immobilized onto a support such as nitrocellulose or a membrane. A mixture comprising at least one protein is first separated on an acrylamide gel, and the separated proteins are then transferred from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized proteins are exposed to at least one antibody with reactivity against at least one antigen of interest. The bound antibodies may be detected by various methods, including the use of radiolabeled antibodies.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with

numerous other mRNA s which encode a multitude of proteins. However, isolated nucleic acid encoding a plant CPA-FAS includes, by way of example, such nucleic acid in cells ordinarily expressing a DES, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (*i.e.*, the oligonucleotide may single-stranded), but may contain both the sense and anti-sense strands (*i.e.*, the oligonucleotide may be double-stranded).

The term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated. The term "purified" or "to purify" also refer to the removal of contaminants from a sample. The removal of contaminating proteins results in an increase in the percent of polypeptide of interest in the sample. In another example, recombinant polypeptides are expressed in plant, bacterial, yeast, or mammalian host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

The term "sample" is used in its broadest sense. In one sense it can refer to a plant cell or tissue. In another sense, it is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

DESCRIPTION OF THE INVENTION

The present invention relates to genes encoding proteins involved in plastid division and morphology, and the encoded proteins, and to methods of use of these genes and proteins.

In particular, the present invention provides compositions comprising isolated Ftn2 (ARC6), ARC5, and Fzo-like genes and polypeptides. The present invention also provides methods for using Ftn2, ARC5, and Fzo-like genes, and polypeptides; such methods include but are not limited to altering plant phenotype by transgenic expression of Ftn2, ARC5, and Fzo-like genes and antisense genes. The description below provides specific, but not limiting, illustrative examples of embodiments of the present invention.

I. Identification of Prokaryotic-Type Plastid Division and Related Genes

Genes involved in plastid division can be identified and characterized by different routes. One route is to identify mutants in plastid division. Such mutants have been identified in Arabidopsis. A set of mutants, referred to as *arc* mutants (for accumulation and replication of chloroplasts), have been isolated and analyzed (Marrison JL et al. (1999) The Plant Journal 18(6): 651-662), the mesophyll chloroplasts differ considerably from wild type in number, size and shape. The *arc* mutant phenotypes are stable and result from single nuclear recessive mutation. Eleven independent nuclear *ARC* genes have been identified so far, and 5 *arc* mutants analyzed with respect to their effects on the stages of the proplastid and chloroplast division processes (Marrison JL et al. (1999) The Plant Journal 18(6): 651-662). These effects are summarized as follows: *ARC1* is involved in the down-regulation of proplastid division, but is in a separate pathway from the other four *ARC* genes, and *arc1* leads to increased proplastid division; *ARC6* is involved in the initiation of both proplastid and chloroplast division, and *arc6* completely suppresses proplastid and chloroplast division, but allows extended expansion until the chloroplasts are about 20 time larger than wild type chloroplasts; *ARC11* is involved in the central positioning of the division constriction, and in *arc11* the constriction is asymmetric; *ARC3* controls chloroplast expansion, and the abnormally rapid expansion of *arc3* chloroplasts prevents chloroplast division; *ARC5* facilitates the separation of the two daughter plastids, and in *arc5* the chloroplasts remain dumb-bell shaped and continue to expand (Marrison JL et al. (1999) The Plant Journal 18(6): 651-662). The map positions of ARC5 (on chromosome 3) and ARC 11 and ARC6 (both on chromosome 5) have also been reported (Marrison JL et al. (1999) The Plant Journal 18(6): 651-662).

However, these plastid division mutants have not yet led to the identification of specific genes involved in plastid division. Another route to identify such genes is based upon homology to genes in other organisms, where the homologs may carry out similar functions in plant plastids. For example, homologs to genes involved in cyanobacterial division may, if present in plants, have a role in plastid division. However, this route depends upon the prior identification of such genes.

The development of the present invention involved first the identification of cyanobacterial genes involved in cell division, then the identification of homologous genes in plants and other cyanobacteria.

A. Cyanobacterial division genes

Cyanobacteria are ancient relatives of chloroplasts and structurally similar to Gram-negative prokaryotes, and perform plant-type photosynthesis. Therefore, it is contemplated that genes present in cyanobacteria which are involved in cell division may have orthologs present in plants which are involved in plastid division.

To date, the genetic control of cell division has been studied much less in cyanobacteria than it has in *Escherichia coli*, *Bacillus subtilis* or *Caulobacter crescentus*. Morphologically aberrant mutants of cyanobacteria presumably impaired in cell division, recovered with high frequency after chemical mutagenesis (Ingram LO and Thurston EL (1970) Protoplasma 71:51-75; Ingram LO and Van Baalen C (1970) J. Bateriol. 102:784-789; Ingram LO, Van Baalen C and Fisher WD (1972) J. Bateriol. 11:614-621; Ingram LO and Fisher W.D.(1973a) J. Bacteriol. 113:995-1005; Ingram LO and Fisher W.D.(1973b) J. Bacteriol. 113:1006-1014; Ingram LO and Blackwell MM (1975) J. Bacteriol. 123:743-746; Zhevner VD, Glazer VM, and Shestakov SV (1973) Mikrobiologiya 42:290-297), were described almost three decades ago. Since that time, little information has been obtained about cyanobacterial genes that are involved in the regulation of cell division. Recently, a cyanobacterial gene that encodes an ortholog of cell division protein FtsZ has been cloned and sequenced from *Anabaena* PCC 7120 and other cyanobacteria (Doherty HM and Adams DG (1995) Gene:93-99; Zhang CC, Huguenin S, and Friry A (1995) Res. Microbiol. 146:445-455). It is contemplated that the discovery of additional cyanobacterial genes involved in cell division and cell differentiation would enhance

understanding of the mechanism and regulation of morphogenesis of both bacteria and plant chloroplasts, and that such genes would be useful to control such processes, for example in bacterial fermenters and in crop and horticultural plants.

In an effort to identify additional genes involved in cell division, transposon mutagenesis, using an improved transposon with an increase in rates of transposition of about two orders of magnitude, was applied to cyanobacteria. Effective transposons have been previously developed, resulting in Tn5 and its improved progeny, for example Tn*5*-1058, where Tn5-1058 and its progeny were characterized by (i) a much stronger promoter driving the antibiotic-resistance operon, (ii) enhanced transposition, and (iii) an *Escherichia coli* origin of replication within the transposon that facilitates recovery of the mutated gene. This vector allows the cloning of sequences contiguous with the transposon, by cutting genomic DNA with a restriction endonuclease that does not cut within the transposon, recircularizing in vitro, and transforming *E. coli* with the resulting ligation mixture (e.g., Black TA, Cai Y, and Wolk CP (1993) Mol. Microbiol. 9:77-84; Cai Y, and Wolk CP (1997) J. Bacteriol. 179:258-266; Ernst A, Black T, Cai Y, Panoff JM, Tiwari DN, and Wolk CP (1992) J. Bacteriol. 174:6025-6032; Wolk CP, Cai Y, and Panoff JM (1991) Proc. Natl. Acad. Sci. USA 88:5355-5359). The transposon subsequently developed by the inventors, Tn*5*-692, represented yet a further improved, demonstrating about a 100-fold increase in the rate of transposition. During the development of the present invention, the use of Tn5-692 provided large numbers of transposon mutants of *Anabaena variabilis* strain ATCC 29413 (PCC 7120) and of *Synechococcus* sp. PCC 7942. Of these transposon-derived mutants, two new cell division mutants of PCC 7942 have now been characterized.

Filamentous cyanobacterial cell division mutants described many years ago showed two distinct phenotypes (Ingram LO, and Fisher WD (1973a) J. Bacteriol. 113:999-1005): septate filaments containing cross-walls, apparently impaired in the terminal stages of cell separation; and serpentine forms that divide sporadically to produce multinucleoidal long cells. The gene mutated in a septate mutant of *Synechococcus* sp. strain PCC 7942 as a consequence of insertional inactivation (Dolganov N, and Grossman AR (1993) J. Bacteriol. 175:7644-7651) was identified and characterized.

By use of transposon mediated mutation, the inventors have discovered mutants of the second, serpentine phenotype. Cells of these mutants, designated FTN2 and FTN6 of *Synechococcus* sp. strain PCC 7942, have the appearance of long filaments that divide occasionally, at variable positions along the cell. Characterization of the protein Ftn2 revealed presence of a DnaJ domain, a (single) tetratricopeptide repeat (TPR) and a leucine zipper motif, which suggest that Ftn2 may function as part of a complex with one or more other proteins and may be regulatory.

DnaJ domains are characteristic of a family of molecular chaperones. Proteins in this family, from bacterial to human, have three distinct domains: (i) a highly conserved J domain of approximately 70 amino acids, often found near the N-terminus, which mediates interaction of DnaJ (a.k.a., Hsp40) with Hsp70 (DnaK) and regulates the ATPase activity of the latter; (ii) a glycine and phenylalanine (G/F)-rich region of unknown function that may act as a flexible linker; and (iii) a cysteine-rich region (C domain) that contains four CXXCXGXG motifs, and resembles a zinc-finger domain (Ohtsuka K, and Hata M (2000) Int. J. Hyperthermia). Although not originally identified as an *fts* gene, *dnaJ* shares with *fts* genes the property that its inactivation leads to a filamentous phenotype (Paciorek J, Kardys K, Lobacz B, and Wolska KI (1997) Acta Microbiol. Pol. 46:7-17). Cheetham and Caplan (Cheetham ME, and Caplan AJ (1998) Cell Stress Chaperones 3:28-36) classified DnaJ/Hsp40 homologs into three groups: type I have all three of these domains; type II have only the J and G/F domains; and type III, like Ftn2, have only a J domain. DnaK proteins are highly versatile chaperones that assist a large variety of processes (Bukau B (1999 ed.) Molecular Chaperones and Folding Catalysts-Regulation, Cellular Function and Mechanisms, Hardwood, Amsterdam; Bukau B, and Horwich AL (1998) Cell 92:351-366; Cai Y, and Wolk CP (1997) J. Bacteriol. 179:258-266; Fink A (1999) Physiological Rev. 79:425-449; Gething MJ (1997) Nature 388:329-331; Hartl FU (1996) Nature 381:571-579), from folding of newly synthesized proteins to facilitation of proteolytic degradation of unstable proteins (Laufen T, Mayer MP, and Heiter P (1995) Sci. USA 96:5452-5457). This functional diversity requires that DnaK proteins associate promiscuously with misfolded proteins or selectively with folded substrates, including with regulatory proteins of low abundance.

The tetratricopeptide repeat (TPR) of, typically, 34 amino acids was first described in the yeast cell division cycle regulator Cdc23p (Sikorski RS, Boguski MS, Goebl M, and Heieter P (1990) Cell 60:307-317) and was later found in many other proteins (Das AK, Cohen PW, and Barford D (1998) EMBO J. 17:1192-1199; Goebl M, and Yanagida M (1991) Trends Biochem. Sci. 16:173-177; Lamb JR, Tugendreich S, and Hieter P (1995) Trends Biochem. Sci. 20:257-259). TPRs are frequently present in tandem arrays of 3-16 copies, although single (as in FTN2) or paired TPRs are also common (; Lamb JR, Tugendreich S, and Hieter P (1995) Trends Biochem. Sci. 20:257-259). Processes involving TPR proteins include cell-cycle control, repression of transcription, response to stress, protein kinase inhibition, mitochondrial and peroxisomal protein transport, and neurogenesis (Goebl M, and Yanagida M (1991) Trends Biochem. Sci. 16:173-177). There appears to be no common biochemical function connecting TRP-containing proteins, although the TRP forms scaffolds that mediate protein-protein interactions and, often, the assembly of multiprotein complexes.

Ftn6 is homologous with hypothetical protein Sll1939 of PCC 6803 (BLAST score, 59; Expect = 10^{-08}). ORF *slr*2041, situated 1325 bp from *sll*1939 on the opposite strand of DNA, predicts a cell-division protein, DivK.

B. Plant Plastid Division and Related Genes

The cyanobacterial *Ftn2* genes and proteins were then used to search for homologous genes from Arabidopsis. Any such genes discovered were then characterized, in order to determine if in fact they are plastid division or related genes. Arabidopsis and cyanobacterial *Ftn2* genes and proteins were then used to search for homologous genes from other cyanobacteria, plants, both vascular and non-vascular, and algae.

The product of the cyanobacterial *Ftn2* gene from *Synechococcus* sp. strain PCC 7942 was discovered to share a similarity with an unknown protein of *Arabidopsis thaliana* (AB016888|Q9FIG9; BLAST score, 72.8; Expect = 1 x 10^{-11}). It was therefore contemplated that this ortholog was involved in plastid division in Arabidopsis cells. The encoded product of this Arabidopsis *Ftn2* ortholog was predicted to posses a chloroplast transit peptide (from a web-based program (http://HypothesisCreator.net/iPSORT/), with the amino acid sequence MEALS HVGIG LSPFQ LCRLP PATTK LRRSH. The Arabidopsis protein was also

predicted to possess a DnaJ domain profile according to ProfileScan (http://www.isrec.isb-sib.ch/software/PFSCAN_form.html), and a Myb DNA-binding domain, according to InterProScan (http://www.ebi.ac.uk/interpro/scan.html).

The inventors subsequently identified, sequenced and characterized the orthologous gene and protein from Arabidopsis (see Figures 1 and 2). Based upon these results, the inventors discovered a novel chloroplast division gene in *Arabidopsis thaliana*; because chloroplast division gene in *Arabidopsis thaliana* is a homologue of the recently identified cell division gene *Ftn2* from a cyanobacterium *Synechococcus*, the Arabidopsis gene is designated *AtFtn2*.

The gene *AtFtn2* is a nuclear gene coding for a chloroplast-targeted protein with an unconventional DnaJ-like N-terminal domain. The inventors further discovered that the *Arabidopsis arc6* mutant, as described above and in which plastid division is completely blocked, and whose cells contain grossly enlarged chloroplasts, carries a point mutation in *AtFtn2* resulting in premature termination of the translated protein. Moreover, the *arc6* mutant phenotype can be rescued by a wild-type copy of *AtFtn2*. In the *arc6* mutant, FtsZ filaments are highly fragmented and disorganized and do not form a ring at mid plastid typical for wild type chloroplasts. Therefore, it is contemplated that AtFtn2 is important for stability and/or assembly of the cytoskeletal plastid-dividing FtsZ protein rings.

The inventors have also discovered Ftn2 homologues in additional cyanobacterial and plant species, but not in completely and partially sequenced genomes of non-cyanobacterial prokaryotes and thus in which Ftn2 homologues appear to be absent.

Therefore, the inventors have discovered a novel gene family involved in plastid and in cyanobacterial prokaryotic division, the *Ftn2* gene family. It is contemplated that *Ftn2* genes and proteins are present in, and thus can be isolated from and/or used in, any organism which possess plastids; such organisms include plants, both vascular and non-vascular, algae, and some parasitic protists which contain vestigial plastids. It is also contemplated that Ftn2 genes and proteins are present in photosynthetic bacteria such as cyanobacteria.

The inventors have discovered additional genes involved in plastid division and/or morphology, *ARC5* and *Fzo-like* genes.

Mutants of *ARC5* exhibit defects in chloroplast constriction, have enlarged, dumbbell-shaped chloroplasts, and are rescued by a wild-type copy of *ARC5*. The *ARC5* gene product shares similarity with the dynamin family of GTPases, which mediate endocytosis, mitochondrial division, and other organellar fission and fusion events in eukaryotes. Phylogenetic analysis showed that ARC5 is related to a group of dynamin-like proteins unique to plants. A green fluorescent protein (GFP)-ARC5 fusion protein localizes to a ring at the chloroplast division site. Chloroplast import and protease protection assays indicate that the ARC5 ring is positioned on the outer surface of the chloroplast. Thus, ARC5 is the first cytosolic component of the chloroplast division complex to be identified. ARC5 has no obvious counterparts in prokaryotes, suggesting that it evolved from a dynamin-related protein present in the eukaryotic ancestor of plants.

Fzo-like genes were discovered by searching the Arabidopsis genomic database using as the query sequence the yeast protein Fzo 1, which in the yeast functions in the control of mitochondrial morphology. The results indicated a related gene in Arabidopsis, referred to as Fzo-like gene, on chromosome 1, At1g03160 on BAC clone F10O3. At least two Arabidopsis lines with T-DNA insertions exhibited abnormalities in chloroplast size and number, indicating the Fzo-like genes functions in plastid division. Knock-out experiments demonstrate that chloroplast development and division are both impaired, where dumbbell-shape chloroplasts with constriction in the middle are frequently observed. Localization experiments with an Fzo-like/GFP fusion protein indicated that the fusion protein is localized to the vesicle-like structures associated with (or near) the chloroplast. The level of AtFzo-like-GFP is positively correlated with the numbers of the vesicle-like structures. Thus, AtFzo-like protein is involved in plastid division and/or morphology.

II. Prokaryotic-Type Division and Related *Ftn2, ARC5, and Fzo-like* Genes and Polypeptides

A. Prokaryotic-Type Division and Related Genes

The present invention provides compositions comprising an isolated nucleic acid sequence comprising prokaryotic-type division and related genes; in particular embodiments, the invention provides compositions comprising isolated *Ftn2, ARC5,* or *Fzo-like* genes. In

some embodiments, the sequences comprise plant *Ftn2, ARC5,* or *Fzo-like* gene; in other embodiments, the sequences comprise Arabidopsis *Ftn2, ARC5,* or *Fzo-like* genes; in other embodiments, the sequences comprise algal *Ftn2, ARC5,* or *Fzo-like* genes; in other embodiments, the sequences comprise cyanobacterial *Ftn2, ARC5,* or *Fzo-like* genes. In different specific embodiments, isolated nucleic acid sequences comprise a nucleic acid sequence as shown in the Figures and/or as described in Table 3, or encode an amino acid sequence as shown in the Figures and/or as described in Table 3.

The present invention also provides compositions comprising an isolated nucleic acid sequence comprising an antisense sequence of prokaryotic-type division and related genes; in particular embodiments, the antisense sequences are directed to *Ftn2, ARC5,* or *Fzo-like* genes. In some embodiments, the sequences comprise an antisense sequence of a plant *Ftn2, ARC5,* or *Fzo-like* gene; in other embodiments, the sequences comprise an antisense sequence of an Arabidopsis *Ftn2, ARC5,* or *Fzo-like* gene; in other embodiments, the sequences comprise an antisense sequence of a cyanobacterial *Ftn2, ARC5,* or *Fzo-like* gene. In different specific embodiments, the sequences comprise antisense sequences of the sequences shown in the Figures and described in Table 3.

The present invention also provides compositions comprising an isolated nucleic acid sequence comprising a sequence encoding any of the *Ftn2, ARC5,* and *Fzo-like* polypeptides as described below, including but not limited to variants, homologs, truncation mutants, and fusion proteins.

B. Prokaryotic-Type Division and Related *Ftn2, ARC5,* and *Fzo-like* Polypeptides

The present invention provides compositions comprising purified prokaryotic-type division and related polypeptides; in particular embodiments, the polypeptides comprise Ftn2, ARC5, or Fzo-like polypeptides, as well as compositions comprising variants, homologs, mutants or fusion proteins thereof. In some embodiments, the polypeptide comprises a plant Ftn2, ARC5, or Fzo-like polypeptide; in other embodiments, the polypeptide comprises an Arabidopsis Ftn2, ARC5, or Fzo-like polypeptide; in other embodiments, the polypeptide comprises an algal Ftn2, ARC5, or Fzo-like polypeptide; in yet other embodiments, the polypeptide comprises a cyanobacterial Ftn2, ARC5, or Fzo-like polypeptides. In different

specific embodiments, the polypeptide is encoded by a nucleic acid sequence as shown in the Figures and/or as described in Tables 3, 10, and 11, or comprises an amino acid sequence as shown in the Figures and/or as described in Tables 3, 10 and 11.

Ftn2, ARC5, and Fzo-like polypeptides are involved in prokaryotic-type division and/or morphology.

In some embodiments, in both photosynthetic prokaryotes and plants, the Ftn2 polypeptide is contemplated to possess a DnaJ domain, a (single) tetratricopeptide repeat (TPR) and a leucine zipper motif, which domains indicate that the Ftn2 functions as part of a complex with one or more other proteins and is a regulatory protein. In plants, the Ftn2 polypeptide is contemplated to further possess an N-terminal plastid targeting sequence, and to be membrane bound. Although it is not necessary to understand the mechanism in order to practice the present invention, and the present invention is not intended to be limited to any particular mechanism or hypothesis, it is hypothesized that the Ftn2 proteins function in regulation of the assembly and stability of the FtsZ plastid dividing ring proteins. This hypothesis is based upon the observations noted above, that in the *arc6* mutants (which lack Ftn2 proteins), little short FtsZ filaments, instead of PD rings, are observed (as described in Example 2).

An Ftn2 polypeptide is a very large protein (in Arabidopsis, it is about 800 to about 830 amino acids long); exemplary but non-limiting sequences are provided in Figs. 2 and 6. An Ftn2 polypeptide can be roughly defined by three regions. The N-terminal contains the DnaJ-like domain, and is exhibits a high degree of homology among Ftn2 proteins obtained from different sources. The large central region is fairly variable, and exhibits a lower degree of homology among the different Ftn2 proteins. The C-terminal is more highly conserved, and therefore exhibits a higher degree of homology. The result is that when considered as a whole, homologous Ftn2 proteins possess about 15% or greater identity or about 38% or greater similarity to AtFtn2 protein. However, the N-terminal and C-terminal regions possess a higher degree of similarity and a higher degree of identity than do the whole proteins.

Thus, in some embodiments, an Ftn2 polypeptide of the present invention comprises at least one of the three regions described above, an N-terminus DnaJ-like domain, a variable central region, and a more conserved C terminal region, and possesses at least some of the

Ftn2 characteristics as described above and in the Examples, where the characteristics include the effects of the absence or decrease in the amount of Ftn2 protein normally occurring in a cell.

In Arabidopsis, a mutation in the Ftn2 gene results in an *arc* (accumulation and replication of chloroplasts) mutant, the *arc6* mutant. The evidence described in Example 2, including the observations that the sequences of Ftn2 from a wild-type background and the sequences of *arc6-1*, *arc6-2*, and *arc6-3*, are essentially the same except that the a C -> T transition at position 1141 in the gene results in a premature stop codon and results in a truncated protein of about 324 amino acids, and that the *arc6* mutant is rescued by a wild-type copy of AtFtn2, indicate that *AtFtn2* gene is *ARC6*.

In some embodiments, ARC5 is also a fairly large protein of almost 800 amino acids; exemplary but non-limiting sequences are provided in Figures 11, 14, 15, and 16. In Arabidopsis, ARC5 exists in two forms, a longer form and a shorter form. The amino acid sequences of ARC5 were deduced from the cDNA sequence; the long form of the cDNA encodes a protein of 777 amino acids and 87.2 kDa, whereas the shorter form of the cDNA encodes a protein of 741 amino acids and 83.5 kDa. In addition, the ARC5 protein contains three motifs found in other dynamin-like proteins: a conserved N-terminal GTPase domain, a pleckstrin homology (PH) domain shown in some proteins to mediate membrane association, and a C-terminal GTPase Effector Domain (GED) thought to interact directly with the GTPase domain and to mediate self-assembly. The shorter cDNA encoded a protein of 741 amino acids and 83.5 kDa identical to that of the larger gene product except for the absence of 36 amino acids encoded by the sequence of the 15th intron.

Thus, in some embodiments, an ARC5 polypeptide of the present invention comprises at least one of the three regions or motifs described above, a conserved N-terminal GTPase domain, a pleckstrin homology (PH) domain, and a C-terminal GTPase Effector Domain (GED), and possesses at least some of the ARC5 characteristics as described above and in the Examples, where the characteristics include the effects of the absence or decrease in the amount of ARC5 protein normally occurring in a cell.

The evidence described in Example 6, which includes the point mutation in At3g19730 /At3g19720 in *arc5*, complementation of the mutant phenotype by the wild-type

gene, and ability of a fragment from At3g19730 /At3g19720 to confer an *arc5*-like phenotype in wild-type plants when expressed in the antisense orientation, indicate that the *ARC5* locus and At3g19730 /At3g19720 represent the same gene. Moreover, in Arabidopsis, the *ARC5* transcript is alternatively spliced. The longer cDNA contained a sequence that was spliced out of the shorter cDNA as the 15th intron; however, its presence in the longer cDNA did not interrupt the reading frame.

In some embodiments, an Fzo-like protein is also fairly large, of slightly more than about 640 amino acids; exemplary but non-limiting sequences are provided in Figures 19 and 22. In Arabidopsis, an Fzo-like of about 642 amino acids has a predicted chloroplast transit peptide, a GTPase domain and two a predicted trans-membrane domains. The evidence described in Example 7 indicates that Fzo-like proteins are involved in plastid division and/or morphology. In some embodiments, An Fzo-like polypeptide

Thus, in some embodiments, an Fzo-like polypeptide of the present invention comprises at least one of the regions described above, chloroplast transit peptide, a GTPase domain and two a predicted trans-membrane domains, and possesses at least some of the Fzo-like characteristics as described above and in the Examples, where the characteristics include the effects of the absence or decrease in the amount of ARC5 protein normally occurring in a cell.

In some embodiments of the present invention, the polypeptide is a purified product, obtained from expression of a native gene in a cell, while in other embodiments it may be a product of chemical synthetic procedures, and in still other embodiments it may be produced by recombinant techniques using a prokaryotic or eukaryotic host (*e.g.*, by bacterial, yeast, higher plant, insect, and mammalian cells in culture). In some embodiments, depending upon the host employed in a recombinant production procedure, the polypeptide of the present invention may be glycosylated or may be non-glycosylated. In other embodiments, the polypeptides of the invention may also include an initial methionine amino acid residue.

In other embodiments, the present invention provides purified Ftn2, ARC5, and Fzo-like peptides encoded by any of the nucleic acid sequences described above and below, where the purified Ftn2, ARC5, and Fzo-like peptides are post-translationally modified. Such modifications include processing, such as by cleavage of peptide fragments. It is

contemplated that newly translated AtFtn2 comprises a plastid peptide sequence, which is cleaved off during import of the protein into the plastid. Thus, AtFtn2 peptides of the present invention include newly translated Ftn2 proteins and post-translationally processed proteins.

Purification of Ftn2, ARC5, and Fzo-like Peptides

In some embodiments of the present invention, Ftn2, ARC5, and Fzo-like polypeptides purified from organisms are provided; such organisms may be transgenic organism, comprising a heterologous Ftn2, ARC5, or Fzo-like gene. The present invention provides purified Ftn2, ARC5, and Fzo-like polypeptides as well as a variant, homolog, mutant or fusion protein thereof, as described elsewhere.

The present invention also provides methods for recovering and purifying Ftn2, ARC5, and Fzo-like polypeptides from an organism; such organisms include single and multi-cellular organisms. Typically, the cells are first disrupted and fractionated before subsequent enzyme purification; disruption and fractionation methods are well-known. Purification methods are also well-known, and include, but are not limited to, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography.

The present invention further provides nucleic acid sequences having a coding sequence of the present invention (*e.g.*, SEQ ID NOs: 1, 11, 14, 19, and 22) fused in frame to a marker sequence that allows for expression alone or both expression and purification of the polypeptide of the present invention. A non-limiting example of a marker sequence is a hexahistidine tag that may be supplied by a vector, for example, a pQE-30 vector which adds a hexahistidine tag to the N terminus of a plastid division and/or morphology polypeptide (e.g., Ftn2, ARC5, and Fzo-like) and which results in expression of the polypeptide in the case of a bacterial host, and more preferably by vector PT-23B, which adds a hexahistidine tag to the C terminal of an plastid division and/or morphology polypeptide and which results in improved ease of purification of the polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a

mammalian host is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson *et al.* (1984) Cell, 37:767).

Chemical Synthesis of Ftn2, ARC5, and Fzo-like Polypeptides

In an alternate embodiment of the invention, the coding sequence of an Ftn2, ARC5, or Fzo-like polypeptide is synthesized, whole or in part, using chemical methods well known in the art (See *e.g.*, Caruthers *et al.* (1980) Nucl. Acids Res. Symp. Ser., 7:215-233; Crea and Horn (1980) Nucl. Acids Res., 9:2331; Matteucci and Caruthers (1980) Tetrahedron Lett., 21:719; and Chow and Kempe (1981) Nucl. Acids Res., 9:2807-2817). In other embodiments of the present invention, the protein itself is produced using chemical methods to synthesize either an entire Ftn2, ARC5, or Fzo-like amino acid sequence or a portion thereof. For example, peptides are synthesized by solid phase techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography (See *e.g.*, Creighton (1983) Proteins Structures And Molecular Principles, W H Freeman and Co, New York N.Y.). In other embodiments of the present invention, the composition of the synthetic peptides is confirmed by amino acid analysis or sequencing (See *e.g.*, Creighton, supra).

Direct peptide synthesis can be performed using various solid-phase techniques (Roberge *et al.* (1995) Science, 269:202-204) and automated synthesis may be achieved, for example, using ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer. Additionally, an amino acid sequence of an Ftn2, ARC5, or Fzo-like polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with other sequences to produce a variant polypeptide.

Generation of Ftn2, ARC5, and Fzo-like Polypeptide Antibodies

In some embodiments of the present invention, antibodies are generated to allow for the detection and characterization of Ftn2, ARC5, and Fzo-like proteins. The antibodies may be prepared using various immunogens. In one embodiment, the immunogen is an Arabidopsis Ftn2, ARC5, or Fzo-like peptide (*e.g.*, an amino acid sequence as depicted in SEQ ID NOs:2, 13, 16, 17, 18, 21, 24, or fragments thereof) to generate antibodies that

recognize Arabidopsis Ftn2, ARC5, and Fzo-like proteins; in another embodiment, the immunogen is a cyanobacterial Ftn2, ARC5, or Fzo-like peptide (*e.g.*, an amino acid sequence as depicted in SEQ ID NO:5, or fragments thereof) to generate antibodies that recognize a cyanobacterial Ftn2, ARC5, or Fzo-like protein. In yet other embodiments, an antibody generated from an immunogenic Ftn2, ARC5, or Fzo-like peptide or fragment recognizes more than one Ftn2, ARC5, or Fzo-like protein or fragment; thus, in these embodiments, the antibodies are cross-reactive. In exemplary embodiments, an antibody prepared against an Arabidopsis Ftn2, ARC5, or Fzo-like peptide or fragment recognizes Ftn2, ARC5, or Fzo-like proteins from both Arabidopsis and cyanobacteria, and an antibody prepared against an cyanobacterial Ftn2, ARC5, or Fzo-like peptide or fragment recognizes Ftn2, ARC5, or Fzo-like proteins from both cyanobacteria and Arabidopsis. Such antibodies include, but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and Fab expression libraries.

Various procedures known in the art may be used for the production of polyclonal antibodies directed against a prokaryotic-type or plastid division and/or morphology gene (e.g., Ftn2, ARC5, or Fzo-like). For the production of antibody, various host animals can be immunized by injection with the peptide corresponding to an Ftn2, ARC5, or Fzo-like epitope including but not limited to rabbits, mice, rats, sheep, goats, etc. In a preferred embodiment, the peptide is conjugated to an immunogenic carrier (*e.g.*, diphtheria toxoid, bovine serum albumin (BSA), or keyhole limpet hemocyanin (KLH)). Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacille Calmette-Guerin) and Corynebacterium parvum).

For preparation of monoclonal antibodies directed toward an Ftn2, ARC5, or Fzo-like peptide, it is contemplated that any technique that provides for the production of antibody molecules by continuous cell lines in culture finds use with the present invention (See *e.g.*, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). These include but are not limited to the hybridoma technique

originally developed by Köhler and Milstein (Köhler and Milstein (1975) Nature, 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (See *e.g.*, Kozbor *et al.* (1983) Immunol. Tod., 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.* (1985) *in* Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

In addition, it is contemplated that techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) find use in producing an Ftn2, ARC5, or Fzo-like peptide-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse *et al.* (1989) Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for an Ftn2, ARC5, or Fzo-like peptide.

It is contemplated that any technique suitable for producing antibody fragments finds use in generating antibody fragments that contain the idiotype (antigen binding region) of the antibody molecule. For example, such fragments include but are not limited to: F(ab')2 fragment that can be produced by pepsin digestion of the antibody molecule; Fab' fragments that can be generated by reducing the disulfide bridges of the F(ab')2 fragment, and Fab fragments that can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, it is contemplated that screening for the desired antibody is accomplished by techniques known in the art (*e.g.*, radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels, for example), Western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many methods are known in the art for detecting binding in an

immunoassay and are within the scope of the present invention. As is well known in the art, the immunogenic peptide should be provided free of the carrier molecule used in any immunization protocol. For example, if the peptide was conjugated to KLH, it may be conjugated to BSA, or used directly, in a screening assay.

In some embodiments of the present invention, the foregoing antibodies are used in methods known in the art relating to the expression of an Ftn2, ARC5, or Fzo-like peptide (*e.g.*, for Western blotting), measuring levels thereof in appropriate biological samples, etc. The antibodies can be used to detect Ftn2, ARC5, and Fzo-like peptides in a biological sample, as for example from a plant or from a cyanobacteria. The biological sample can be an extract of a tissue or cells, or a sample fixed for microscopic examination.

The biological samples are then be tested directly for the presence of an Ftn2, ARC5, or Fzo-like peptide using an appropriate strategy (*e.g.*, ELISA or radioimmunoassay) and format (*e.g.*, microwells, dipstick (*e.g.*, as described in International Patent Publication WO 93/03367), etc. Alternatively, proteins in the sample can be size separated (*e.g.*, by polyacrylamide gel electrophoresis (PAGE), in the presence or not of sodium dodecyl sulfate (SDS), and the presence of an Ftn2, ARC5, or Fzo-like peptide detected by immunoblotting (Western blotting). Immunoblotting techniques are generally more effective with antibodies generated against a peptide corresponding to an epitope of a protein, and hence, are particularly suited to the present invention.

III. Methods of Identifying Ftn2, ARC5, and Fzo-like Genes and Related Genes

Some embodiments of the present invention contemplate methods to isolate nucleic acid sequences encoding a prokaryotic-type or plastid division and/or morphology protein (e.g., Ftn2, ARC5, and Fzo-like proteins). In some embodiments, the methods involve first preparation of a cDNA library from an appropriate source, for example tissue or cells in which prokaryotic-type division occurs, such as in cyanobacteria or plants. The methods involve next subtracting highly abundant sequences from the library, sequencing the remaining library clones, and comparing the encoded amino acid sequences to the amino acid sequence of either cyanobacterial Ftn2 (for example, SEQ ID NO: 5) or Arabidopsis Ftn2, ARC5, or Fzo-like (egg, SEQ ID NO:2, 13, 16, 17, 18, 21, and 24) to select putative Ftn2,

ARC5, or Fzo-like peptide candidate ESTs. The methods involve next assembling a clone encoding a complete putative Ftn2, ARC5, or Fzo-like peptide, and characterizing the expression products of such sequences so discovered. Alternatively, the methods involve first an examination of an expressed sequence tag (EST) database from an appropriate source, for example tissue or cells in which prokaryotic-type division occurs, such as in cyanobacteria or plants, in order to discover novel potential Ftn2, ARC5, or Fzo-like encoding sequences. These methods next involve sequencing likely candidate sequences, and characterizing the expression products of such sequences so discovered.

Employing these methods resulted in the discovery of an Arabidopsis *Ftn2*, as described in illustrative Examples. The isolated novel coding sequence was demonstrated to encode an Ftn2, as described in the illustrative Examples. These methods were also used to discover other homologous Ftn2, ARC5, and Fzo-like genes, coding sequences, or ESTs from other plants, including vascular plant, and non-vascular plants such as mosses and ferns, and other cyanobacteria, as shown in Example 3, 6, and 7 (see Tables 3, 10, and 11). It is contemplated that these methods can also be used to discover other homologous Ftn2, ARC5, and Fzo-like genes, coding sequences, or ESTs from other plants, both vascular and non-vascular, algae, and other cyanobacteria. It is also contemplated that homologous Ftn2, ARC5, and Fzo-like genes are present in parasitic protists, which are unicellular eukaryotes containing vestigial plastids. These protists are sensitive to the herbicide ROUND-UP, and possess biosynthetic and metabolic pathways which are characteristic of plant plastids, although the protist plastid genome appears to be reduced compared to plant plastid genomes. Exemplary protists include but are not limited to the malarial protist *Plasmodium falciarum* and *Toxoplasma gondii*.

The Ftn2, ARC5, and Fzo-like coding sequences described above can be used to locate and isolate Ftn2, ARC5, and Fzo-like genes, by methods well known in the art. In some methods to isolate the gene, a ^{32}P-radiolabeled Ftn2, ARC5, or Fzo-like coding sequence (or cDNA) from a particular source is used to screen, by DNA-DNA hybridization, a genomic or cDNA library constructed from the source genomic DNA. Single isolated clones that test positive for hybridization are proposed to contain part or all of the plastid division and/or morphology gene, and are sequenced. The sequence of a positive cloned Ftn2, ARC5, or Fzo-

like genomic DNA is used to confirm the identity of the gene as an Ftn2, ARC5, or Fzo-like gene. If a particular clone encodes only part of the gene, additional clones that test positive for hybridization to an Ftn2, ARC5, or Fzo-like coding sequence (or cDNA) are isolated and sequenced. Comparison of the full-length sequence of the Ftn2, ARC5, or Fzo-like gene to the cDNA are used to determine the location of introns, if they are present.

Other methods for identifying other Ftn2, ARC5, or Fzo-like genes are also known. Such methods include utilizing structural predictions used to find related proteins. For example, protein motifs may be used to search for identical or similar proteins present in various databases, as well as their coding sequences (as described further below). Hydropathy profiles can also be used to search databases for similar protein profiles. In yet other methods, cross-hybridizing by Southern blot analysis can be used to screen libraries, and the hybridizing DNA sequenced.

IV. Additional Plastid Division and Related Genes

The present invention provides isolated nucleic acid sequences encoding a prokaryotic-type or plastid division and/or morphology gene (e.g., Ftn2, ARC5, or Fzo-like genes). For example, some embodiments of the present invention provide isolated polynucleotide sequences that are capable of hybridizing to Ftn2, ARC5, and Fzo-like coding sequences (for example, SEQ ID NOs: 1, 3, 4, 11, 12, 14, 15, 19, 20, 22, and 23) under conditions of low to high stringency as long as the polynucleotide sequence capable of hybridizing encodes a protein that retains a desired biological activity of the naturally occurring Ftn2, ARC5, or Fzo-like. In preferred embodiments, hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex and confer a defined "stringency" as explained above (See *e.g.*, Wahl *et al.* (1987) Meth. Enzymol., 152:399-407, incorporated herein by reference).

In other embodiments, an isolated nucleic acid sequence encoding an Ftn2, ARC5, or Fzo-like peptide which is homologous to an Ftn2, ARC5, or Fzo-like as described in the Examples (for example, SEQ ID NOs; 2, 5, 13, 16, 17, 18, 21, and 24) is provided; in some embodiments, the sequence is obtained from a plant or cyanobacteria.

In other embodiments of the present invention, alleles of an Ftn2, ARC5, or Fzo-like gene are provided. In preferred embodiments, alleles result from a mutation, (*i.e.*, a change in the nucleic acid sequence) and generally produce altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to deletions, additions or substitutions of nucleic acids. Each of these types of changes may occur alone, or in combination with the others, and at the rate of one or more times in a given sequence.

In other embodiments of the present invention, the polynucleotide sequence encoding an Ftn2, ARC5, or Fzo-like gene is extended utilizing the nucleotide sequences *(e.g.*, SEQ ID NOs:3, 11, 14, 19, and 22) in various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, it is contemplated that polymerase chain reaction (PCR) finds use in the present invention. This is a direct method that uses universal primers to retrieve unknown sequence adjacent to a known locus (Gobinda *et al.* (1993) PCR Methods Applic., 2:318-322). First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

In another embodiment, inverse PCR is used to amplify or extend sequences using divergent primers based on a known region (Triglia *et al.* (1988) Nucleic Acids Res., 16:8186). The primers may be designed using Oligo 4.0 (National Biosciences Inc, Plymouth Minn.), or another appropriate program, to be, for example, 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72 °C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template. In yet another embodiment of the present invention, capture PCR (Lagerstrom *et al.* (1991) PCR Methods Applic., 1:111-119) is used. This is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome (YAC) DNA. Capture PCR also requires multiple restriction enzyme

digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR. In still other embodiments, walking PCR is utilized. Walking PCR is a method for targeted gene walking that permits retrieval of unknown sequence (Parker *et al.* (1991) Nucleic Acids Res., 19:3055-60). The PROMOTERFINDER kit (Clontech) uses PCR, nested primers and special libraries to "walk in" genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions. In yet other embodiments of the present invention, add TAIL PCR is used as a preferred method for obtaining flanking genomic regions, including regulatory regions (Lui and Whittier, (1995); Lui *et al.* (1995)).

Preferred libraries for screening for full length cDNAs include libraries that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred, in that they contain more sequences that contain the 5' and upstream gene regions. A randomly primed library may be particularly useful in cases where an oligo d(T) library does not yield full-length cDNA. Genomic Libraries are useful for obtaining introns and extending 5' sequence.

In yet other embodiments, databases containing complete or partial maps of a source genome can be utilized; exemplary genomes are described in Example 1. The flanking sequences can then be obtained from the database once an Ftn2, ARC5, or Fzo-like gene is identified from the source.

V. Variant Plastid Division Peptides

In some embodiments, the present invention provides isolated variants of the disclosed nucleic acid sequence encoding plastid division and/or morphology (e.g., Ftn2, ARC5, and Fzo-like) peptides, and the polypeptides encoded thereby; the peptide variants include mutants, fragments, fusion proteins or functional equivalents of Ftn2, ARC5, and Fzo-like peptides. Thus, nucleotide sequences of the present invention are engineered in order to alter an Ftn2, ARC5, or Fzo-like peptide coding sequence for a variety of reasons, including but not limited to alterations that modify the cloning, processing and/or expression of the gene product (such alterations include inserting new restriction sites, altering glycosylation patterns, and changing codon preference) as well as varying the regulatory and/or enzymatic

activity (such changes include but are not limited to differing substrate affinities, differing substrate preferences and utilization, differing inhibitor affinities or effectiveness, differing reaction kinetics, varying subcellular localization, and varying protein processing and/or stability).

Mutants of an Ftn2, ARC5, or Fzo-like peptide

Some embodiments of the present invention provide mutant forms of an Ftn2, ARC5, or Fzo-like peptide (*i.e.*, muteins). In preferred embodiments, variants result from mutation, (*i.e.*, a change in the nucleic acid sequence) and generally produce altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene may have none, one, or many mutant forms. Common mutational changes that give rise to variants are generally ascribed to deletions, additions or substitutions of nucleic acids. Each of these types of changes may occur alone, or in combination with the others, and at the rate of one or more times in a given sequence.

It is contemplated that is possible to modify the structure of a peptide having an activity (*e.g.*, a prokaryotic-type or plastid division and morphology activity) for such purposes as altering the activity of the peptide. Such modified peptides are considered functional equivalents of peptides having an activity of an Ftn2, ARC5, or Fzo-like peptide as defined herein. A modified peptide can be produced in which the nucleotide sequence encoding the polypeptide has been altered, such as by substitution, deletion, or addition. In some embodiments, these modifications do not significantly reduce the synthetic activity of the modified enzyme. In other words, construct "X" can be evaluated in order to determine whether it is a member of the genus of modified or variant Ftn2, ARC5, and Fzo-like peptides of the present invention as defined functionally, rather than structurally. In some embodiments, the activity of variant Ftn2, ARC5, and Fzo-like peptides is evaluated by the methods described in Examples 2 or 6. For example, a variant Ftn2 can be evaluated in an *arc6* mutant, as described in Example 2; an expressed functional Ftn2 peptide will partially or completely restore the mutant to a wild-type phenotype. Accordingly, in some embodiments the present invention provides nucleic acids encoding an Ftn2, ARC5, or Fzo-like peptide that

complement the coding region of an Ftn2, ARC5, or Fzo-like coding sequence provided herein (for example, SEQ ID NOs: 1, 3, 4, 11, 14, 19, or 22).

As described above, mutant forms of Ftn2, ARC5, and Fzo-like peptides are also contemplated as being equivalent to those peptides and DNA molecules that are set forth in more detail herein. For example, it is contemplated that isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (*i.e.*, conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Accordingly, some embodiments of the present invention provide variants of Ftn2, ARC5, and Fzo-like peptides disclosed herein containing conservative replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic (aspartate, glutamate); (2) basic (lysine, arginine, histidine); (3) nonpolar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan); and (4) uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine). Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic (aspartate, glutamate); (2) basic (lysine, arginine, histidine), (3) aliphatic (glycine, alanine, valine, leucine, isoleucine, serine, threonine), with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic (phenylalanine, tyrosine, tryptophan); (5) amide (asparagine, glutamine); and (6) sulfur -containing (cysteine and methionine) (*e.*g., Stryer ed. (1981) *Biochemistry*, pg. 17-21, 2nd ed, WH Freeman and Co.). Whether a change in the amino acid sequence of a peptide results in a functional homolog can be readily determined by assessing the ability of the variant peptide to function in a fashion similar to the wild-type protein. Peptides having more than one replacement can readily be tested in the same manner.

More rarely, a variant includes "nonconservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Analogous minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs (*e.g.*, LASERGENE software, DNASTAR Inc., Madison, Wis.).

Mutants of Ftn2, ARC5, and Fzo-like peptides can be generated by any suitable method well known in the art, including but not limited to site-directed mutagenesis, randomized "point" mutagenesis, and domain-swap mutagenesis in which portions of the Sterculia CPA-FAS cDNA are "swapped" with the analogous portion of other plant or bacterial CPA-FAS-encoding cDNAs (Back and Chappell (1996) PNAS 93: 6841-6845).

Variants may be produced by methods such as directed evolution or other techniques for producing combinatorial libraries of variants. Thus, the present invention further contemplates a method of generating sets of combinatorial mutants of the present Ftn2, ARC5, and Fzo-like proteins, as well as truncation mutants, and is especially useful for identifying potential variant sequences (*i.e.*, homologs) that possess the biological activity of a Ftn2, ARC5, or Fzo-like (*e.g.*, role in prokaryotic-type cell or plastid division and/or morphology). In addition, screening such combinatorial libraries is used to generate, for example, novel Ftn2, ARC5, or Fzo-like homologs that possess novel substrate specificities or other biological activities.

It is contemplated that Ftn2, ARC5, and Fzo-like coding nucleic acids (*e.g.*, SEQ ID NOs: 1, 3, 4, 11, 14, 19, and 22 and fragments and variants thereof) can be utilized as starting nucleic acids for directed evolution. These techniques can be utilized to develop Ftn2, ARC5, or Fzo-like peptide variants having desirable properties such as increased synthetic activity or altered affinity.

In some embodiments, artificial evolution is performed by random mutagenesis (*e.g.*, by utilizing error-prone PCR to introduce random mutations into a given coding sequence). This method requires that the frequency of mutation be finely tuned. As a general rule, beneficial mutations are rare, while deleterious mutations are common. This is because the combination of a deleterious mutation and a beneficial mutation often results in an inactive enzyme. The ideal number of base substitutions for targeted gene is usually between 1.5 and 5 (Moore and Arnold (1996) Nat. Biotech., 14, 458-67; Leung *et al.* (1989) Technique, 1:11-15; Eckert and Kunkel (1991) PCR Methods Appl., 1:17-24; Caldwell and Joyce (1992) PCR Methods Appl., 2:28-33; and Zhao and Arnold (1997) Nuc. Acids. Res., 25:1307-08). After mutagenesis, the resulting clones are selected for desirable activity (*e.g.*, role in prokaryotic-type cell division, as described in Example 2). Successive rounds of mutagenesis and

selection are often necessary to develop enzymes with desirable properties. It should be noted that only the useful mutations are carried over to the next round of mutagenesis.

In other embodiments of the present invention, the polynucleotides of the present invention are used in gene shuffling or sexual PCR procedures (*e.g.*, Smith (1994) Nature, 370:324-25; U.S. Pat. Nos. 5,837,458; 5,830,721; 5,811,238; 5,733,731). Gene shuffling involves random fragmentation of several mutant DNAs followed by their reassembly by PCR into full length molecules. Examples of various gene shuffling procedures include, but are not limited to, assembly following DNase treatment, the staggered extension process (STEP), and random priming *in vitro* recombination. In the DNase mediated method, DNA segments isolated from a pool of positive mutants are cleaved into random fragments with DNaseI and subjected to multiple rounds of PCR with no added primer. The lengths of random fragments approach that of the uncleaved segment as the PCR cycles proceed, resulting in mutations in present in different clones becoming mixed and accumulating in some of the resulting sequences. Multiple cycles of selection and shuffling have led to the functional enhancement of several enzymes (Stemmer (1994) Nature, 370:398-91; Stemmer (1994) Proc. Natl. Acad. Sci. USA, 91, 10747-10751; Crameri *et al.* (1996) Nat. Biotech., 14:315-319; Zhang *et al.* (1997) Proc. Natl. Acad. Sci. USA, 94:4504-09; and Crameri *et al.* (1997) Nat. Biotech., 15:436-38). Variants produced by directed evolution can be screened for function in prokaryotic-type or plastid division and/or morphology by the methods described subsequently (see Example 2).

Homologs

Still other embodiments of the present invention provide isolated nucleic acid sequence encoding Ftn2, ARC5, and Fzo-like homologs, and the polypeptides encoded thereby. Some homologs of Ftn2, ARC5, and Fzo-like peptides have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered proteins are rendered either more stable or less stable to proteolytic degradation or other cellular process that result in destruction of, or otherwise inactivate plant CPA-FAS. Such homologs, and the genes that encode them, can be utilized to alter the activity of Ftn2, ARC5, and Fzo-like peptides by modulating the half-life of the protein. For instance, a short half-life

can give rise to more Ftn2, ARC5, or Fzo-like peptide biological effects. Other homologs have characteristics that are either similar to wild-type Ftn2, ARC5, or Fzo-like peptides, or which differ in one or more respects from wild-type Ftn2, ARC5, or Fzo-like peptides.

The amino acid sequences of plant and cyanobacterial Ftn2 proteins were searched for protein motifs. One motif is a putative DnaJ domain (AtFtn2 residues 89-153; Scc_ PCC 7942_Ftn2 residues 6-70) as determined by the InterProScan program (InterPro accession IPR001623, Pfam conserved domain pfam00226). However, ClustalW alignment of this domain with all predicted DnaJ domains from the Pfam database (277 sequences) revealed that the central HPD motif essential for DnaJ proteins is not present in AtFtn2 or other plant and cyanobacterial ftn2 homologues (see Figure 4).

Another domain discovered through a Pfam-HMM search in the plant Ftn2 proteins is a putative myb domain (residues 677-690, see Figures 3 and 5), albeit with low expectation value (0.63). Sequence alignment with entries from the Prosite database indicated that this motif represents only about a half of a typical myb domain.

Yet another domain in AtFtn2 is from one to three transmembrane domains; various software tools predicted up to three putative transmembrane helices (Table 2).

The Scc_ PCC 7942_Ftn2 also possesses a single TPR repeat (residues 136-169) as determined by the InterProScan program, and a leucine zipper pattern (residues 234-255) as determined by the Prosite-Protein against PROSITE program (http://ca.expasy.org/tools/scnpsite.html/).

Accordingly, in some embodiments, the present invention provides an Ftn2 prokaryotic-type division peptide comprising at least the DnaJ-like domain (where the DnaJ-like domain is missing the central HPD amino acid (histidine-proline-aspartate), AtFtn2 residues 89-153; Scc_ PCC 7942_Ftn2 residues 6-70), or the nucleic acid sequences corresponding thereto. In yet other embodiments of the present invention, it is contemplated that nucleic acid sequences suspected of encoding an Ftn2 homolog is screened by comparing motifs. In some embodiments, the deduced amino acid sequence can be analyzed for the presence of the DnaJ-like amino acid motif (AtFtn2 residues 89-153; Scc_ PCC 7942_Ftn2 residues 6-70), the putative myb domain (AtFtn2 residues 677-690), TPR repeat

(Scc_PCC7942_Ftn2 residues 136-169) or a leucine zipper pattern (Scc_PCC7942_Ftn2 residues 234-255).

In some embodiments of the combinatorial mutagenesis approach of the present invention, the amino acid sequences for a population of prokaryotic-type or plastid division and/or morphology peptides (e.g., Ftn2, ARC5, or Fzo-like) homologs are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, Ftn2, ARC5, and Fzo-like homologs from one or more species, or Ftn2, ARC5, and Fzo-like homologs from the same species but which differ due to mutation. Amino acids that appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences.

In a preferred embodiment of the present invention, the combinatorial Ftn2, ARC5, or Fzo-like library is produced by way of a degenerate library of genes encoding a library of polypeptides that each include at least a portion of candidate Ftn2, ARC5, or Fzo-like -protein sequences. For example, a mixture of synthetic oligonucleotides is enzymatically ligated into gene sequences such that the degenerate set of candidate Ftn2, ARC5, or Fzo-like sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of Ftn2, ARC5, or Fzo-like sequences therein.

There are many ways by which the library of potential Ftn2, ARC5, or Fzo-like homologs can be generated from a degenerate oligonucleotide sequence. In some embodiments, chemical synthesis of a degenerate gene sequence is carried out in an automatic DNA synthesizer, and the synthetic genes are ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential Ftn2, ARC5, or Fzo-like sequences. The synthesis of degenerate oligonucleotides is well known in the art (See *e.g.*, Narang (1983) Tetrahedron Lett., 39:3-9; Itakura *et al.* (1981) Recombinant DNA, in Walton (ed.), Proceedings of the 3rd Cleveland Symposium on Macromolecules, Elsevier, Amsterdam, pp 273-289; Itakura *et al.* (1984) Annu. Rev. Biochem., 53:323; Itakura *et al.* (1984) Science 198:1056; Ike *et al.* (1983) Nucl. Acid Res., 11:477). Such techniques have been employed in the directed evolution of other proteins (See *e.g.*, Scott *et al.* (1980) Science, 249:386-390; Roberts *et al.* (1992) Proc. Natl. Acad. Sci. USA, 89:2429-2433; Devlin *et al.* (1990) Science, 249: 404-

406; Cwirla *et al.* (1990) Proc. Natl. Acad. Sci. USA, 87: 6378-6382; as well as U.S. Pat. Nos. 5,223,409, 5,198,346, and 5,096,815).

Truncation Mutants of Ftn2, ARC5, or Fzo-like Proteins

In addition, the present invention provides isolated nucleic acid sequences encoding fragments of Ftn2, ARC5, or Fzo-like (*i.e.*, truncation mutants), and the polypeptides encoded by such nucleic acid sequences. In preferred embodiments, the Ftn2, ARC5, or Fzo-like fragment is biologically active.

In some embodiments of the present invention, when expression of a portion of an Ftn2, ARC5, or Fzo-like protein is desired, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli* (Ben-Bassat *et al.* (1987) J. Bacteriol., 169:751-757) and Salmonella typhimurium and its *in vitro* activity has been demonstrated on recombinant proteins (Miller *et al.* (1990) Proc. Natl. Acad. Sci. USA, 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing such recombinant polypeptides in a host that produces MAP (*e.g.*, *E. coli* or CM89 or S. cerevisiae), or *in vitro* by use of purified MAP.

Fusion Proteins Containing Ftn2, ARC5, or Fzo-like Proteins

The present invention also provides nucleic acid sequences encoding fusion proteins incorporating all or part of Ftn2, ARC5, or Fzo-like proteins, and the polypeptides encoded by such nucleic acid sequences. In some embodiments, the fusion proteins have an Ftn2, ARC5, or Fzo-like functional domain with a fusion partner. Accordingly, in some embodiments of the present invention, the coding sequences for the polypeptide (*e.g.*, an Ftn2, ARC5, or Fzo-like functional domain) is incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. In one embodiment, a single fusion product polypeptide comprises an Ftn2, ARC5, or Fzo-like peptide fused to a marker protein; in some embodiments, the marker protein is GFP.

In some embodiments of the present invention, chimeric constructs code for fusion proteins containing a portion of an Ftn2, ARC5, or Fzo-like protein and a portion of another gene. In some embodiments, a fusion protein has biological activity similar to the wild type Ftn2, ARC5, or Fzo-like protein (*e.g.*, have at least one desired biological activity of an Ftn2, ARC5, or Fzo-like protein). In other embodiments, the fusion protein has altered biological activity.

In other embodiments of the present invention, chimeric constructs code for fusion proteins containing an *Ftn2, ARC5, or Fzo-like* gene or portion thereof and a leader or other signal sequences which direct the protein to targeted subcellular locations. Such sequences are well known in the art, and direct proteins to locations such as the chloroplast, the mitochondria, the endoplasmic reticulum, the tonoplast, the golgi network, and the plasmalemma.

In addition to utilizing fusion proteins to alter biological activity, it is widely appreciated that fusion proteins can also facilitate the expression and/or purification of proteins, such as an Ftn2, ARC5, or Fzo-like protein of the present invention. Accordingly, in some embodiments of the present invention, an Ftn2, ARC5, or Fzo-like protein is generated as a glutathione-S-transferase (*i.e.*, GST fusion protein). It is contemplated that such GST fusion proteins enables easy purification of an Ftn2, such as by the use of glutathione-derivatized matrices (*See e.g.*, Ausabel *et al.* (eds.) (1991) Current Protocols in Molecular Biology, John Wiley & Sons, NY).

In another embodiment of the present invention, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of an Ftn2, ARC5, or Fzo-like protein allows purification of the expressed Ftn2, ARC5, or Fzo-like fusion protein by affinity chromatography using a Ni^{2+} metal resin. In still another embodiment of the present invention, the purification leader sequence is then subsequently removed by treatment with enterokinase (*See e.g.*, Hochuli *et al.* (1987) J. Chromatogr., 411:177; and Janknecht *et al.* Proc. Natl. Acad. Sci. USA, 88:8972). In yet other embodiments of the present invention, a fusion gene coding for a purification sequence appended to either the N (amino) or the C (carboxy) terminus allows for

affinity purification; one example is addition of a hexahistidine tag to the carboxy terminus of an Ftn2, ARC5, or Fzo-like protein which was optimal for affinity purification.

Techniques for making fusion genes are well known. Essentially, the joining of various nucleic acid fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment of the present invention, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, in other embodiments of the present invention, PCR amplification of gene fragments is carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed to generate a chimeric gene sequence (*See e.g.*, Current Protocols in Molecular Biology, supra).

Screening Gene Products

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, and for screening cDNA libraries for gene products having a certain property. Such techniques are generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of Ftn2 homologs. The most widely used techniques for screening large gene libraries typically comprise cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate sequences created by combinatorial mutagenesis techniques.

Accordingly, in some embodiment of the present invention, candidate *Ftn2, ARC5, or Fzo-like* gene products are displayed on the surface of a cell or viral particle, and the product detected by any of several methods. In other embodiments of the present invention, the gene library is cloned into the gene for a surface membrane protein of a bacterial cell, and the

resulting fusion protein detected by panning (WO 88/06630; Fuchs *et al.* (1991) BioTechnol., 9:1370-1371; and Goward *et al.* (1992) TIBS 18:136-140). In other embodiments of the present invention, fluorescently labeled molecules that bind an Ftn2, ARC5, or Fzo-like peptide can be used to score for potentially functional Ftn2, ARC5, or Fzo-like homologs. Cells are visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In an alternate embodiment of the present invention, the gene library is expressed as a fusion protein on the surface of a viral particle. For example, foreign peptide sequences are expressed on the surface of infectious phage in the filamentous phage system, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd, and fl are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (See *e.g.*, WO 90/02909; WO 92/09690; Marks *et al.* (1992) J. Biol. Chem., 267:16007-16010; Griffths *et al.* (1993) EMBO J., 12:725-734; Clackson *et al.* (1991) Nature, 352:624-628; and Barbas *et al.* (1992) Proc. Natl. Acad. Sci., 89:4457-4461).

In another embodiment of the present invention, the recombinant phage antibody system (*e.g.*, RPAS, Pharmacia Catalog number 27-9400-01) is modified for use in expressing and screening of Ftn2, ARC5, or Fzo-like combinatorial libraries. The pCANTAB 5 phagemid of the RPAS kit contains the gene that encodes the phage gIII coat protein. In some embodiments of the present invention, the Ftn2, ARC5, or Fzo-like combinatorial gene library is cloned into the phagemid adjacent to the gIII signal sequence such that it is expressed as a gIII fusion protein. In other embodiments of the present invention, the phagemid is used to transform competent *E. coli* TG1 cells after ligation. In still other embodiments of the present invention, transformed cells are subsequently infected with M13KO7 helper phage to rescue the phagemid and its candidate *Ftn2, ARC5, or Fzo-like* gene insert. The resulting recombinant phage contain phagemid DNA encoding a specific candidate Ftn2, ARC5, or

Fzo-like protein and display one or more copies of the corresponding fusion coat protein. In some embodiments of the present invention, the phage-displayed candidate proteins that are capable of, for example, interacting with other prokaryotic-type proteins, are selected or enriched by panning. The bound phage is then isolated, and if the recombinant phage express at least one copy of the wild type gIII coat protein, they will retain their ability to infect *E. coli*. Thus, successive rounds of reinfection of *E. coli* and panning will greatly enrich for Ftn2, ARC5, or Fzo-like homologs, which can then be screened for further biological activities.

In light of the present disclosure, other forms of mutagenesis generally applicable will be apparent to those skilled in the art in addition to the aforementioned rational mutagenesis based on conserved versus non-conserved residues. For example, Ftn2, ARC5, or Fzo-like homologs can be generated and screened using, for example, alanine scanning mutagenesis and the like (Ruf *et al.* (1994) Biochem., 33:1565-1572; Wang *et al.* (1994) J. Biol. Chem., 269:3095-3099; Balint (1993) Gene 137:109-118; Grodberg *et al.* (1993) Eur. J. Biochem., 218:597-601; Nagashima *et al.* (1993) J. Biol. Chem., 268:2888-2892; Lowman *et al.* (1991) Biochem., 30:10832-10838; and Cunningham *et al.* (1989) Science, 244:1081-1085), by linker scanning mutagenesis (Gustin *et al.* (1993) Virol., 193:653-660; Brown *et al.* (1992) Mol. Cell. Biol., 12:2644-2652; McKnight *et al.* Science, 232:316); or by saturation mutagenesis (Meyers *et al.* (1986) Science, 232:613).

VI. Expression of Cloned Plastid Division and Related Genes

In other embodiment of the present invention, nucleic acid sequences corresponding to plastid division and/or morphology (*e.g., Ftn2, ARC5, or Fzo-like*) genes, homologs and mutants as described above may be used to generate recombinant DNA molecules that direct the expression of the encoded protein product in appropriate host cells.

As will be understood by those of skill in the art, it may be advantageous to produce Ftn2, ARC5, or Fzo-like -encoding nucleotide sequences possessing non-naturally occurring codons. Therefore, in some preferred embodiments, codons preferred by a particular prokaryotic or eukaryotic host (Murray *et al.* (1989) Nucl. Acids Res., 17) can be selected, for example, to increase the rate of Ftn2, ARC5, or Fzo-like expression or to produce

recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

A. Vectors for Production of Plastid Division and Related Proteins

The nucleic acid sequences of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the nucleic acid sequence may be included in any one of a variety of expression vectors for expressing a polypeptide. In some embodiments of the present invention, vectors include, but are not limited to, chromosomal, nonchromosomal and synthetic DNA sequences (*e.g.*, derivatives of SV40, bacterial plasmids, phage DNA; baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, and viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies). It is contemplated that any vector may be used as long as it is replicable and viable in the host.

In particular, some embodiments of the present invention provide recombinant constructs comprising one or more of the nucleic sequences as broadly described above (*e.g.*, SEQ ID NOs: 1, 3, 4, 11, 14, 19, and 22). In some embodiments of the present invention, the constructs comprise a vector, such as a plasmid or viral vector, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In preferred embodiments of the present invention, the appropriate nucleic acid sequence is inserted into the vector using any of a variety of procedures. In general, the nucleic acid sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art.

Large numbers of suitable vectors are known to those of skill in the art, and are commercially available. Such vectors include, but are not limited to, the following vectors: 1) Bacterial -- pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, phagescript, psiX174, pbluescript SK, pBSKS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); and 2) Eukaryotic -- pWLNEO, pSV2CAT, pOG44, PXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Any other plasmid or vector may be used as long as they are replicable and viable in the host. In some preferred embodiments of the present invention, plant expression vectors comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome

binding sites, polyadenylation sites, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. In other embodiments, DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

In certain embodiments of the present invention, a nucleic acid sequence of the present invention within an expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. Promoters useful in the present invention include, but are not limited to, the LTR or SV40 promoter, the *E. coli* lac or trp, the phage lambda P_L and P_R, T3 and T7 promoters, and the cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, and mouse metallothionein-I promoters and other promoters known to control expression of gene in prokaryotic or eukaryotic cells or their viruses. In other embodiments of the present invention, recombinant expression vectors include origins of replication and selectable markers permitting transformation of the host cell (*e.g.*, dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or tetracycline or ampicillin resistance in *E. coli*).

In some embodiments of the present invention, transcription of the DNA encoding polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Enhancers useful in the present invention include, but are not limited to, the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

In other embodiments, the expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. In still other embodiments of the present invention, the vector may also include appropriate sequences for amplifying expression.

B. Host Cells for Production of Plastid Division and Related Polypeptides

In a further embodiment, the present invention provides host cells comprising any of the above-described constructs. In some embodiments of the present invention, the host cell is a higher eukaryotic cell (*e.g.*, a plant cell). In other embodiments of the present invention,

the host cell is a lower eukaryotic cell (*e.g.*, a yeast cell). In still other embodiments of the present invention, the host cell can be a prokaryotic cell (*e.g.*, a bacterial cell). Specific examples of host cells include, but are not limited to, *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, as well as *Saccharomycees cerivisiae*, *Schizosaccharomycees pombe*, *Drosophila* S2 cells, Spodoptera Sf9 cells, Chinese hamster ovary (CHO) cells, COS-7 lines of monkey kidney fibroblasts, (Gluzman (1981) Cell 23:175), 293T, C127, 3T3, HeLa and BHK cell lines, NT-1 (tobacco cell culture line), root cell and cultured roots in rhizosecretion (Gleba *et al.* (1999) Proc Natl Acad Sci USA 96: 5973-5977). Other examples include microspore-derived cultures of oilseed rape. (Weselake RJ and Taylor DC (1999) Prog. Lipid Res. 38: 401), and transformation of pollen and microspore culture systems. Yet other examples include red and green algal cells. Further examples are described in the Examples.

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by any of the recombinant sequences of the present invention described above. In some embodiments, introduction of the construct into the host cell can be accomplished by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (See *e.g.*, Davis *et al.* (1986) Basic Methods in Molecular Biology). Alternatively, in some embodiments of the present invention, a polypeptide of the invention can be synthetically produced by conventional peptide synthesizers.

Proteins can be expressed in eukaryotic cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from a DNA construct of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.* (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y.

In some embodiments of the present invention, following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. In other embodiments of the present invention, cells are

typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. In still other embodiments of the present invention, microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

C. Transgenic Plants, Seeds, and Plant Parts

In other embodiments, the present invention provides plants, seeds, plant cells and/or plant parts comprising any of the above-described constructs. Plants are transformed with a heterologous gene encoding an Ftn2, ARC5, or Fzo-like protein or transformed with a fusion gene encoding a fusion polypeptide expressing an Ftn2, ARC5, or Fzo-like protein according to procedures well known in the art. It is contemplated that the heterologous genes are utilized to alter the level of the proteins encoded by the heterologous genes. It is further contemplated that the heterologous genes are utilized to change the phenotype of the transgenic plants; such changes in phenotype are contemplated to include but not be limited to change in plastid size, number per cell, and shape.

Plants

The methods of the present invention are not limited to any particular plant. Indeed, a variety of plants are contemplated in different embodiments, including but not limited to tomato, potato, tobacco, pepper, rice, corn, barley, wheat, *Brassica*, *Arabidopsis*, sunflower, soybean, poplar, and pine. In some embodiments, plants include oil-producing species, which are plant species that produce and store triacylglycerol in specific organs, primarily in seeds; fatty acids are synthesized in the plastid. Such species include but are not limited to soybean (*Glycine max*), rapeseed and canola (including *Brassica napus* and *B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*), coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*) and peanut (*Arachis hypogaea*). The group also includes non-agronomic species which are useful in developing appropriate expression vectors such as tobacco, rapid cycling *Brassica* species, and

Arabidopsis thaliana, and wild species which may be a source of genes encoding metabolites synthesized in the plastid. Other plants include plants that synthesize desirable compounds in the plastid, such as production of carotenoid pigments, as for example in tomatoes and marigolds, and production of starch, as for example in corn and potatoes.

Vectors

The methods of the present invention contemplate the use of a heterologous gene encoding an Ftn2, ARC5, or Fzo-like polypeptide, as described above. Such genes include any of the sequences described above, including variants and fragments.

Heterologous genes intended for expression in plants are first assembled in expression cassettes comprising a promoter. Methods that are well known to those skilled in the art may be used to construct expression vectors containing a heterologous gene and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are widely described in the art (See *e.g.*, Sambrook. *et al.* (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. *et al.* (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y).

In general, these vectors comprise a nucleic acid sequence of the invention encoding an Ftn2, ARC5, or Fzo-like polypeptide (as described above) operably linked to a promoter and other regulatory sequences (*e.g.*, enhancers, polyadenylation signals, etc.) required for expression in a plant.

Promoters include but are not limited to constitutive promoters, tissue-, organ-, and developmentally-specific promoters, and inducible promoters. Examples of promoters include but are not limited to: constitutive promoter 35S of cauliflower mosaic virus; a wound-inducible promoter from tomato, leucine amino peptidase ("LAP," Chao *et al.* (1999) Plant Physiol 120: 979-992); a chemically-inducible promoter from tobacco, Pathogenesis-Related 1 (PR1) (induced by salicylic acid and BTH (benzothiadiazole-7-carbothioic acid S-methyl ester)); a tomato proteinase inhibitor II promoter (PIN2) or LAP promoter (both inducible with methyl jasmonate); a heat shock promoter (US Pat 5,187,267); a tetracycline-

inducible promoter (US Pat 5,057,422); and seed-specific promoters, such as those for seed storage proteins (*e.g.*, phaseolin, napin, oleosin, and a promoter for soybean beta conglycin (Beachy *et al.* (1985) EMBO J. 4: 3047-3053)). All references cited herein are incorporated in their entirety.

The expression cassettes may further comprise any sequences required for expression of mRNA. Such sequences include, but are not limited to transcription terminators, enhancers such as introns, viral sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments.

A variety of transcriptional terminators are available for use in expression of sequences using the promoters of the present invention. Transcriptional terminators are responsible for the termination of transcription beyond the transcript and its correct polyadenylation. Appropriate transcriptional terminators and those which are known to function in plants include, but are not limited to, the CaMV 35S terminator, the tml terminator, the pea rbcS E9 terminator, and the nopaline and octopine synthase terminator (See *e.g.*, Odell *et al.* (1985) Nature 313:810; Rosenberg *et al.* (1987) Gene, 56:125; Guerineau *et al.* (1991) Mol. Gen. Genet., 262:141; Proudfoot (1991) Cell, 64:671; Sanfacon *et al.* Genes Dev., 5:141 ; Mogen *et al.* (1990) Plant Cell, 2:1261; Munroe *et al.* (1990) Gene, 91:151; Ballad *et al.* (1989) Nucleic Acids Res. 17:7891; Joshi *et al.* (1987) Nucleic Acid Res., 15:9627).

In addition, in some embodiments, constructs for expression of the gene of interest include one or more of sequences found to enhance gene expression from within the transcriptional unit. These sequences can be used in conjunction with the nucleic acid sequence of interest to increase expression in plants. Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize Adh1 gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells (Calais *et al.* (1987) Genes Develop. 1: 1183). Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

In some embodiments of the present invention, the construct for expression of the nucleic acid sequence of interest also includes a regulator such as a nuclear localization signal

(Calderone *et al.* (1984) Cell 39:499; Lassoer *et al.* (1991) Plant Molecular Biology 17:229), a plant translational consensus sequence (Joshi (1987) Nucleic Acids Research 15:6643), an intron (Luehrsen and Walbot (1991) Mol. Gen. Genet. 225:81), and the like, operably linked to the nucleic acid sequence encoding plant CPA-FAS.

In preparing the construct comprising a nucleic acid sequence encoding an Ftn2, ARC5, or Fzo-like polypeptide, various DNA fragments can be manipulated, so as to provide for the DNA sequences in the desired orientation (*e.g.*, sense or antisense) orientation and, as appropriate, in the desired reading frame. For example, adapters or linkers can be employed to join the DNA fragments or other manipulations can be used to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resection, ligation, or the like is preferably employed, where insertions, deletions or substitutions (*e.g.*, transitions and transversions) are involved.

Numerous transformation vectors are available for plant transformation. The selection of a vector for use will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers are preferred. Selection markers used routinely in transformation include the nptII gene which confers resistance to kanamycin and related antibiotics (Messing and Vierra (1982) Gene 19: 259; Bevan *et al.* (1983) Nature 304:184), the bar gene which confers resistance to the herbicide phosphinothricin (White *et al.* (1990) Nucl Acids Res. 18:1062; Spencer *et al.* (1990) Theor. Appl. Genet. 79: 625), the hph gene which confers resistance to the antibiotic hygromycin (Blochlinger and Diggelmann (1984) Mol. Cell. Biol. 4:2929), and the dhfr gene, which confers resistance to methotrexate (Bourouis *et al.* (1983) EMBO J., 2:1099).

In some preferred embodiments, the vector is adapted for use in an *Agrobacterium* mediated transfection process (*See e.g.*, U.S. Pat. Nos. 5,981,839; 6,051,757; 5,981,840; 5,824,877; and 4,940,838; all of which are incorporated herein by reference). Construction of recombinant Ti and Ri plasmids in general follows methods typically used with the more common bacterial vectors, such as pBR322. Additional use can be made of accessory genetic elements sometimes found with the native plasmids and sometimes constructed from foreign

sequences. These may include but are not limited to structural genes for antibiotic resistance as selection genes.

There are two systems of recombinant Ti and Ri plasmid vector systems now in use. The first system is called the "cointegrate" system. In this system, the shuttle vector containing the gene of interest is inserted by genetic recombination into a non-oncogenic Ti plasmid that contains both the cis-acting and trans-acting elements required for plant transformation as, for example, in the pMLJ1 shuttle vector and the non-oncogenic Ti plasmid pGV3850. The second system is called the "binary" system in which two plasmids are used; the gene of interest is inserted into a shuttle vector containing the cis-acting elements required for plant transformation. The other necessary functions are provided in trans by the non-oncogenic Ti plasmid as exemplified by the pBIN19 shuttle vector and the non-oncogenic Ti plasmid PAL4404. Some of these vectors are commercially available.

In other embodiments of the invention, the nucleic acid sequence of interest is targeted to a particular locus on the plant genome. Site-directed integration of the nucleic acid sequence of interest into the plant cell genome may be achieved by, for example, homologous recombination using *Agrobacterium*-derived sequences. Generally, plant cells are incubated with a strain of *Agrobacterium* which contains a targeting vector in which sequences that are homologous to a DNA sequence inside the target locus are flanked by *Agrobacterium* transfer-DNA (T-DNA) sequences, as previously described (U.S. Pat. No. 5,501,967). One of skill in the art knows that homologous recombination may be achieved using targeting vectors which contain sequences that are homologous to any part of the targeted plant gene, whether belonging to the regulatory elements of the gene, or the coding regions of the gene. Homologous recombination may be achieved at any region of a plant gene so long as the nucleic acid sequence of regions flanking the site to be targeted is known.

In yet other embodiments, the nucleic acids of the present invention is utilized to construct vectors derived from plant (+) RNA viruses (*e.g.*, brome mosaic virus, tobacco mosaic virus, alfalfa mosaic virus, cucumber mosaic virus, tomato mosaic virus, and combinations and hybrids thereof). Generally, the inserted plant CPA-FAS polynucleotide of the present invention can be expressed from these vectors as a fusion protein (*e.g.*, coat protein fusion protein) or from its own subgenomic promoter or other promoter. Methods for

the construction and use of such viruses are described in U.S. Pat. Nos. 5,846,795; 5,500,360; 5,173,410; 5,965,794; 5,977,438; and 5,866,785, all of which are incorporated herein by reference.

In some embodiments of the present invention, where the nucleic acid sequence of interest is introduced directly into a plant. One vector useful for direct gene transfer techniques in combination with selection by the herbicide Basta (or phosphinothricin) is a modified version of the plasmid pCIB246, with a CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator (WO 93/07278).

Transformation Techniques

Once a nucleic acid sequence encoding an Ftn2, ARC5, or Fzo-like polypeptide is operatively linked to an appropriate promoter and inserted into a suitable vector for the particular transformation technique utilized (*e.g.*, one of the vectors described above), the recombinant DNA described above can be introduced into the plant cell in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of plant targeted for transformation. In some embodiments, the vector is maintained episomally. In other embodiments, the vector is integrated into the genome.

In some embodiments, direct transformation in the plastid genome is used to introduce the vector into the plant cell (See *e.g.*, U.S. Patent Nos 5,451,513; 5,545,817; 5,545,818; PCT application WO 95/16783); these techniques also result in plastid transformation. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the nucleic acid encoding the RNA sequences of interest into a suitable target tissue (*e.g.*, using biolistics or protoplast transformation with calcium chloride or PEG). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab *et al.* (1990) PNAS, 87:8526; Staub and Maliga, (1992) Plant Cell, 4:39). The presence of cloning sites between these markers allowed creation of a plastid targeting vector introduction of foreign DNA

molecules (Staub and Maliga (1993) EMBO J., 12:601). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial aadA gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'-adenyltransferase (Svab and Maliga (1993) PNAS, 90:913). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the present invention. Plants homoplasmic for plastid genomes containing the two nucleic acid sequences separated by a promoter of the present invention are obtained, and are preferentially capable of high expression of the RNAs encoded by the DNA molecule.

In other embodiments, vectors useful in the practice of the present invention are microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA (Crossway (1985) Mol. Gen. Genet, 202:179). In still other embodiments, the vector is transferred into the plant cell by using polyethylene glycol (Krens *et al.* (1982) Nature, 296:72; Crossway *et al.* (1986) BioTechniques, 4:320); fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies (Fraley *et al.* (1982) Proc. Natl. Acad. Sci., USA, 79:1859); protoplast transformation (EP 0 292 435); direct gene transfer (Paszkowski *et al.* (1984) EMBO J., 3:2717; Hayashimoto *et al.* (1990) Plant Physiol. 93:857).

In still further embodiments, the vector may also be introduced into the plant cells by electroporation. (Fromm, *et al.* (1985) Pro. Natl Acad. Sci. USA 82:5824; Riggs *et al.* (1986) Proc. Natl. Acad. Sci. USA 83:5602). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

In yet other embodiments, the vector is introduced through ballistic particle acceleration using devices (*e.g.*, available from Agracetus, Inc., Madison, Wis. and Dupont, Inc., Wilmington, Del). (See *e.g.*, U.S. Pat. No. 4,945,050; and McCabe *et al.* (1988) Biotechnology 6:923). See also, Weissinger *et al.* (1988) Annual Rev. Genet. 22:421; Sanford *et al.* (1987) Particulate Science and Technology, 5:27 (onion); Svab *et al.* (1990) Proc. Natl. Acad. Sci. USA, 87:8526 (tobacco chloroplast); Christou *et al.* (1988) Plant Physiol., 87:671

(soybean); McCabe *et al.* (1988) Bio/Technology 6:923 (soybean); Klein *et al.* (1988) Proc. Natl. Acad. Sci. USA, 85:4305 (maize); Klein *et al.* (1988) Bio/Technology, 6:559 (maize); Klein *et al.* (1988) Plant Physiol., 91:4404 (maize); Fromm *et al.* (1990) Bio/Technology, 8:833; and Gordon-Kamm *et al.* (1990) Plant Cell, 2:603 (maize); Koziel *et al.* (1993) Biotechnology, 11:194 (maize); Hill *et al.* (1995) Euphytica, 85:119 and Koziel *et al.* (1996) Annals of the New York Academy of Sciences 792:164; Shimamoto *et al.* (1989) Nature 338: 274 (rice); Christou *et al.* (1991) Biotechnology, 9:957 (rice); Datta *et al.* (1990) Bio/Technology 8:736 (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil *et al.* (1993) Biotechnology, 11: 1553 (wheat); Weeks *et al.* (1993) Plant Physiol., 102: 1077 (wheat); Wan *et al.* (1994) Plant Physiol. 104: 37 (barley); Jahne *et al.* (1994) Theor. Appl. Genet. 89:525 (barley); Knudsen and Muller (1991) Planta, 185:330 (barley); Umbeck *et al.* (1987) Bio/Technology 5: 263 (cotton); Casas *et al.* (1993) Proc. Natl. Acad. Sci. USA 90:11212 (sorghum); Somers *et al.* (1992) Bio/Technology 10:1589 (oat); Torbert *et al.* (1995) Plant Cell Reports, 14:635 (oat); Weeks *et al.* (1993) Plant Physiol., 102:1077 (wheat); Chang *et al.*, WO 94/13822 (wheat) and Nehra *et al.* (1994) The Plant Journal, 5:285 (wheat).

In addition to direct transformation, in some embodiments, the vectors comprising a nucleic acid sequence encoding an Ftn2, ARC5, or Fzo-like polypeptide of the present invention are transferred using Agrobacterium-mediated transformation (Hinchee *et al.* (1988) Biotechnology, 6:915; Ishida *et al.* (1996) Nature Biotechnology 14:745). Agrobacterium is a representative genus of the gram-negative family Rhizobiaceae. Its species are responsible for plant tumors such as crown gall and hairy root disease. In the dedifferentiated tissue characteristic of the tumors, amino acid derivatives known as opines are produced and catabolized. The bacterial genes responsible for expression of opines are a convenient source of control elements for chimeric expression cassettes. Heterologous genetic sequences (*e.g.*, nucleic acid sequences operatively linked to a promoter of the present invention), can be introduced into appropriate plant cells, by means of the Ti plasmid of Agrobacterium tumefaciens. The Ti plasmid is transmitted to plant cells on infection by Agrobacterium tumefaciens, and is stably integrated into the plant genome (Schell (1987) Science, 237: 1176). Species which are susceptible infection by Agrobacterium may be transformed *in*

vitro. Alternatively, plants may be transformed *in vivo*, such as by transformation of a whole plant by Agrobacteria infiltration of adult plants, as in a "floral dip" method (Bechtold N, Ellis J, Pelletier G (1993) Cr. Acad. Sci. III - Vie 316: 1194-1199).

Regeneration

After selecting for transformed plant material that can express the heterologous gene encoding a plastid division and/or morphology polypeptide (e.g., Ftn2, ARC5, or Fzo-like polypeptide), whole plants are regenerated. Plant regeneration from cultured protoplasts is described in Evans *et al.* (1983) Handbook of Plant Cell Cultures, Vol. 1: (MacMillan Publishing Co. New York); and Vasil I. R. (ed.), Cell Culture and Somatic Cell Genetics of Plants, Acad. Press, Orlando, Vol. I (1984), and Vol. III (1986). It is known that many plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables, and monocots (*e.g.*, the plants described above). Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted.

Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate and form mature plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. The reproducibility of regeneration depends on the control of these variables.

Generation of Transgenic lines

Transgenic lines are established from transgenic plants by tissue culture propagation. The presence of nucleic acid sequences encoding exogenous Ftn2, ARC5, or Fzo-like polypeptides of the present invention (including mutants or variants thereof) may be transferred to related varieties by traditional plant breeding techniques.

These transgenic lines are then utilized for evaluation of plastid division and/or morphology and agronomic traits. Evaluation of plastid division and/or morphology includes examination of plastid size, number, and shape in the transgenic lines, and comparison to these characteristics in wild-type parent lines. A difference of at least about 10%, preferably of at least about 25%, and more preferably of at least about 50%, from these characteristics in wild-type plants, is indicative of homologous plastid division and/or morphology gene activity in the transgenic lines.

VII. Manipulation of Ftn2, ARC5, and Fzo-like Levels and Function in Plants

Altering the expression of Ftn2, ARC5, or Fzo-like or homologues in crop species via genetic engineering using antisense, RNAi, cosuppression, or overexpression strategies, introducing Ftn2, ARC5, or Fzo-like homologues from plants, algae or cyanobacteria into plants, algae, or cyanobacteria, is contemplated to result in changes in plastid size, shape and/or number. Such changes are contemplated to occur in all types of plastids including chloroplasts, chromoplasts, leucoplasts and amyloplasts, and in all organs including leaves, roots, stems, petals, and seeds depending on the specificity of the promoters used in the construction of the transgenes.

Alterations in plastid size, shape and/or number via genetic engineering of Ftn2, ARC5, or Fzo-like expression in agronomically or horticulturally important plant and algal species is contemplated to result in improved productivity and/or increased vigor due to enhanced photosynthetic capacity, and/or to allow enhanced production of commercially important compounds that accumulate in plastids either naturally or as a result of genetic engineering. Examples of compounds that naturally accumulate in plastids include vitamin E, pro-vitamin A, essential (aromatic) amino acids, pigments (carotenes, xanthophylls, chlorophylls), starch, and lipids. Plants with altered plastid size or number have further applications in improving the efficiency of plastid transformation technologies that are used for the introduction of transgenes into the plastid genome.

It is contemplated, therefore, that the nucleic acids encoding an Ftn2, ARC5, or Fzo-like polypeptide of the present invention may be utilized to either increase or decrease the level of Ftn2, ARC5, or Fzo-like mRNA and/or protein in transfected cells as compared to the

levels in wild-type cells. Such transgenic cells have great utility, including but not limited to further research as to the effects of the overexpression of Ftn2, ARC5, or Fzo-like, and as to the effects as to the underexpression or lack of Ftn2, ARC5, or Fzo-like genes. In particular embodiments, the cells are plant cells.

Accordingly, in some embodiments, expression in plants by the methods described above leads to the overexpression of Ftn2, ARC5, or Fzo-like genes in transgenic plants, plant tissues, plant cells, or seeds.

In other embodiments of the present invention, Ftn2, ARC5, or Fzo-like encoding polynucleotides are utilized to decrease the level of Ftn2, ARC5, or Fzo-like mRNA and/or protein in transgenic plants, plant tissues, plant cells, or seeds as compared to wild-type plants, plant tissues, plant cells, or seeds. One method of reducing Ftn2, ARC5, or Fzo-like expression utilizes expression of antisense transcripts. Antisense RNA has been used to inhibit plant target genes in a tissue-specific manner (*e.g.*, van der Krol *et al.* (1988) Biotechniques 6:958-976). Antisense inhibition has been shown using the entire cDNA sequence as well as a partial cDNA sequence (*e.g.*, Sheehy *et al.* (1988) Proc. Natl. Acad. Sci. USA 85:8805-8809; Cannon *et al.* (1990) Plant Mol. Biol. 15:39-47). There is also evidence that 3' non-coding sequence fragment and 5' coding sequence fragments, containing as few as 41 base-pairs of a 1.87 kb cDNA, can play important roles in antisense inhibition (Ch'ng *et al.* (1989) Proc. Natl. Acad. Sci. USA 86:10006-10010).

Accordingly, in some embodiments, an Ftn2, ARC5, or Fzo-like encoding-nucleic acid of the present invention (*e.g.*, SEQ ID NOs: 1 3, 11, 14, 19, and 22 and fragments and variants thereof) are oriented in a vector and expressed so as to produce antisense transcripts. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The expression cassette is then transformed into plants and the antisense strand of RNA is produced. The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

Furthermore, for antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of the target gene or genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, Solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff, *et al.* (1988) Nature 334:585-591. Ribozymes targeted to the mRNA of a lipid biosynthetic gene, resulting in a heritable increase of the target enzyme substrate, have also been described (Merlo AO et al. (1998) Plant Cell 10: 1603-1621).

Another method of reducing Ftn2, ARC5, or Fzo-like expression utilizes the phenomenon of cosuppression or gene silencing (*See e.g.*, U.S. Pat. No. 6,063,947, incorporated herein by reference). The phenomenon of cosuppression has also been used to

inhibit plant target genes in a tissue-specific manner. Cosuppression of an endogenous gene using a full-length
cDNA sequence as well as a partial cDNA sequence (730 bp of a 1770 bp cDNA) are known (*e.g.*, Napoli *et al.* (1990) Plant Cell 2:279-289; van der Krol *et al.* (1990) Plant Cell 2:291-299; Smith *et al.* (1990) Mol. Gen. Genetics 224:477-481). Accordingly, in some embodiments the nucleic acid sequences encoding an Ftn2, ARC5, or Fzo-like polypeptide of the present invention (*e.g.* including SEQ ID NOs 1, 3, 11, 14, 19, and 22 and fragments and variants thereof) are expressed in another species of plant to effect cosuppression of a homologous gene.

Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For cosuppression, the introduced sequence in the expression cassette, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

An effective method to down regulate a gene is by hairpin RNA constructs. Guidance to the design of such constructs for efficient, effective and high throughput gene silencing have been described (Wesley SV *et al.* (2001) Plant J. 27: 581-590).

VIII. Herbicide Targets

In some embodiments, the plastid division and/or morphology genes of the present invention find use as herbicide targets. The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, it is contemplated that, based on the fact that *ARC6* is found in plants and cyanobacteria but not in animals, fungi or other eukaryotes, the gene product has use as an herbicide target.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosures which follow, the following abbreviations apply: N (normal); M (molar); mM (millimolar); µM (micromolar); mol (moles); mmol (millimoles); µmol (micromoles); nmol (nanomoles); pmol (picomoles); g (grams); mg (milligrams); µg (micrograms); ng (nanograms); l or L (liters); ml (milliliters); µl (microliters); cm (centimeters); mm (millimeters); µm (micrometers); nm (nanometers); °C (degrees Centigrade); WT (wild type); nt (nucleotide(s)); na (nucleic acid(s)); aa (amino acid(s)); *arc* (accumulation and replication of chloroplasts; refers to mutations observed in Arabidopsis which exhibition abnormal chloroplast accumulation and/or replication)

EXAMPLES

The following examples describe the identification and characterization of several Ftn2 coding sequences and encoded amino acid sequences from cyanobacteria and plants, both vascular and non-vascular. A cyanobacterial cell division gene *Ftn2* (accession AF421196) was isolated from *Synechococcus* sp. WH8102 (as described in Examples 4 and 5). The product of this *Ftn2* gene was then discovered to be similar to an unknown protein of *Arabidopsis thaliana*, as well as to predicted products of ORFs from an *Anabaena* strain, a *Nostoc punctiforme*, and a presumptive gene from a *Synechocystis* strain. The *Arabidopsis Ftn2*

gene, which encodes a protein similar to the *Synechococcus* Ftn2 protein, was then isolated, sequenced, and characterized (as described in Examples 1 and 2). The two encoded Ftn2 protein products were then used to discover other Ftn2 encoding nucleic acid and amino acid sequences from other plants and cyanobacteria (as described in Example 3).

EXAMPLE 1

Materials and Methods Utilized to Identify and Characterize *Ftn2* genes

This example describes the materials and methods used to identify and characterize Ftn2 genes in plants and other cyanobacteria.

Gene and protein names

The cyanobacterial cell division gene *Ftn2* (accession AF421196) was isolated from *Synechococcus* sp. WH8102 as described below (and in Koksharova and (2002) J Bacterial: in press in preparation). Although the initial designation of this gene as *Ftn2* conflicts with existing records for ferritin type 2 protein gene *Ftn2* (e.g., accession AJ306614), in this description the designation *Ftn2* refers to the cyanobacterial cell division gene and its plant homologues. Because the Ftn2 plant homologue was isolated and identified in Arabidopsis *arc6* mutant (as described in Example 2 below), the *ARC6* gene (and ARC6 protein) designations may be used. These denote the same entities as AtFtn2 gene and AtFtn2 protein, respectively.

For clarity, the species abbreviation is used as the first part of the name: *AtFtn2* (*Arabidopsis thaliana*), *StFtn2* (*Solanum tuberosum,* potato), *ZmFtn2* (*Zea mays,* maize), *OsFtn2* (*Oryza sativa,* rice), *Nostoc_Ftn2* (*Nostoc punctiforme* ATCC 29133), *MtFtn2* (*Medicago truncatula*), *Pm_MED4_Ftn2* (*Protochlorococcus marinus* MED4), *Pm_MIT9313_Ftn2* (*Protochlorococcus marinus* MIT 9313), *Scc_WH8102_Ftn2* (*Synechococcus* WH8102), *Syn_PCC6803_Ftn2* (*Synechocystis* PCC6803, NP_441990), and *Anabena_Ftn2* (*Anabena* PCC 7120). The DNA and/or protein accession numbers are listed in Table 3 in Example 3 below.

Plant material

The wild type (WT) Arabidopsis thaliana, ecotype Wassiljevskija (Ws), transgenic plants expressing AtFtsZ1-1 or AtFtsZ2-1 antisense constructs (Osteryoung et al.(1998) Plant Cell. 10:1991-2004), AtFtsZ1-1 sense constructs (Stokes et al., 2000) and AtFtsZ2-1-cmyc sense constructs (Vitha et al.(2001) J. Cell. Biol.153:111-119) (all in ecotype Columbia Col-0 background), the Arabidopsis chloroplast division mutants *arc6-1*, *arc6-2* and *arc6-3* (Ws-2 background) and *arc3* (Landsberg erecta background) were grown for five weeks in a growth chamber as described previously (Osteryoung et al.(1998) Plant Cell. 10:1991-2004).

Amplification and sequencing of AtFtn2

Genomic DNA was isolated from WT and *arc6-11*, *arc6-2* and *arc6-3* young leaf tissue using the Plant DNAzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The *AtFtn2* genomic fragment was amplified with the *PfuTurbo* DNA polymerase (Stratagene, La Jolla, CA) using the primers
5' TGTCCAAATTTTATGTGACACTCC 3' (forward) and
5' TTGTGAAAGGCTTGAATGTAAGA 3' (reverse). The amplification product of ~3.8 kb contained the whole *AtFtn2* coding sequence flanked by a 0.5 kb 5' and a 0.2 kb 3' regions. The amplified product was cloned into a *Sma*I-digested pBluescript vector (Startagene). For each plant genotype, DNA isolation, PCR amplification, and cloning of the product were carried out independently for three individual plants to minimize amplification errors. The resulting plasmid DNA was then pooled for each genotype and sequenced in both directions. Sequencer reads were processed, assembled into contigs, and viewed using Phrap, Phred and Consed (see the Software Tools section).

Complementation of the *arc6-1* mutant

The PCR-amplified genomic fragment containing *AtFtn2* (see above) was cloned into a *Sma*I site of a pBJ97 shuttle vector, excised with *Not*I and inserted into a plant transformation vector pMLBART (both vectors obtained from Karl Gordon, CSIRO, Canberra, Australia via John Bowman, University of California, Davis), a derivative of pART27 (Gleave, 1992), that confers resistance to the herbicide glufosinate as a selectable

marker. *Agrobacterium*-mediated transformation of WT and *arc6-1* plants and selection of the glufosinate-resistant T1 plants were performed as described previously (Vitha et al., 2001).

Microscopy

Chloroplast phenotypes were assessed in tips from fully expanded leaves of four week old plants as described previously (Osteryoung et al.(1998) Plant Cell. 10:1991-2004). Cells containing 1-4 chloroplasts were scored as having severe plastid phenotype. The intermediate phenotype was characterized by 10-30 chloroplasts per cell, while cells containing 50 or more chloroplasts were scored as having WT-like phenotype. Images were recorded with Nikon Coolpix 995 (Nikon Corporation, Tokyo, Japan) digital camera.

Immunoblotting and Immunofluorescence of AtFtsZ

Immunoblotting with leaf tissue extracts and immunofluorescence microscopy of leaf mesophyll chloroplasts were performed as previously described (Stokes et al. (2000) Arabidopsis Plant Physiol. 124:1668-1677; Vitha et al.(2001) J. Cell. Biol.153:111-119) using rabbit antipeptide antibodies specific to AtFtsZ1 and AtFtsZ2 (antibodies were designated 1-1A and 2-1A, respectively). For immunofluorescence labeling, a goat anti-rabbit Oregon Green 488 conjugate (Molecular Probes, Eugene, OR) was used at 1:200 dilution. Specimens were viewed with Olympus BH-2 and Leica DMR A2 microscopes equipped with epifluorescence illumination, 100x oil immersion objectives, FITC fluorescence filter sets (excitation 455-495 nm, emission 512-575 nm) and CCD cameras Optronics (Goleta, CA) DEI 750 and Qimaging (Burnaby, B.C., Canada) Retiga 1350ex, respectively. The images were taken either as a single optical section or as a stack of images with spacing 0.5 μm between slices. Image stacks were processed and projected (Brightest Point method) with ImageJ ver. 1.27 software (http://rsb.info.nih.gov/ij/) and further adjusted and cropped using Adobe Photoshop 6.0 (Adobe Systems Inc., San Jose, CA).

Databases and Software Tools

DNA and protein sequence databases were searched with tblastn and blastn (Altschul et al. (1990) J. Mol Biol. 215:403-10) at National Center for Biotechnology Information (NCBI; at http://www.ncbi.nlm.nih.gov), and in the *Arabidopsis thaliana* database at Munich Information Center for Protein Sequences (MIPS; at http://mips.gsf.de/proj/thal/db/index.html). Preliminary sequence data for *Synechococcus* sp. strain WH8102, strain MED4, *Protochlorococcus marinus* strain MT9313 and *Nostoc punctiforme* strain ATCC 29133 were obtained from the DOE Joint Genome Institute (JGI) (at http://www.jgi.doe.gov/JGI_microbial/html/index.html). The *Anabena* sp. PCC 7120 sequence was obtained from the Kazusa DNA Research Institute, Japan (at http://www.kazusa.or.jp/cyano/). The preliminary *Synechococcus* sp. PCC 7002 sequence was obtained from NCBI through a tblastn search of microbial genomes (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/genom_table_cgi).

For predictions of subcellular protein targeting, TargetP ver. 1.01 (Emanuelsson et al. (2000) J. Mol Biol. 300:1005-16) (at http://www.cbs.dtu.dk/services/TargetP/) and Predotar ver. 0.5 (at http://www.inra.fr/Internet/Produits/Predotar/) were used. Prediction of transmembrane domain was performed with HMMTOP ver. 2.0 (Tusnady and Simon (1998) J. Mol Biol. 283:489-506; Tusnady and Simon (2001) Bioinformatics 17:849-50) (at http://www.enzim.hu/hmmtop/), TMHMM ver. 2.0 (Krogh et al. (2001) J. mol Biol. 305:567-580) (at http://www.cbs.dtu.dk/services/TMHMM-2.0/), DAS (Cserzo et al. (1997) Pro t Eng. 10:673-676) (at http://www.sbc.su.se/~miklos/DAS/), SOSUI (Hirokawa et al. (1998) Bioinformatics 14:378-379(at http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0E.html), Split (Juretic et al. (2002) J. Chem Inf Comp Sci: in press) (at http://pref.etfos.hr/split-4.0/); TMPRED (Hofmann and Stoffel (1993) Biol Chem Hoppe-Seyler 374:166) (at http://www.ch.embnet.org/software/TMPRED_form.html) and TopPred2 (Claros and von Heijne (1994) Comput Appl Biosci 10:685-686) (at http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html). Identification of conserved domains was facilitated by searches in the ProDom Protein domain database (Corpet et al. (2000) Nucleic Acids Res. 28:267-9) (at http://prodes.toulouse.inra.fr/prodom/doc/prodom.html) and

through the Conserved Domain Database and Search Service, v1.54 at NCBI (at http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). The PredictProtein service (at http://www.embl-heidelberg.de/predictprotein/predictprotein.html) was further used as interface to access multiple tools for the primary and secondary structure analysis.

The exon/intron prediction for the rice *Ftn2* homologue from the genomic DNA sequence combined results from several algorithms: GeneScan (Burge and Karlin (1997) J Mol Biol. 215:403-10) (at http://genes.mit.edu/GENSCAN.html), GrailEXP v3.3 (Xu and Uberbacher (1997) J Compt Biol. 4:325-38) (at http://compbio.ornl.gov/grailexp/), FGENESH 1.1 (at http://genomic.sanger.ac.uk/gf/gf.shtml) and Genie (Kulp et al. (1996) Proc Int Conf Intell Syst Mol Biol. 4:134-42) (at http://www.fruitfly.org/seq_tools/genie.html). The exon/intron predictions were then compared to the available rice ESTs and to the homology regions with the *Arabidospis* AtFtn2 identified in tblastn search. Sequence manipulation, multiple alignments and shading of aligned sequences were performed using BioEdit 5.09 (at http://www.mbio.ncsu.edu/BioEdit/bioedit.html). DNA sequencing reads were processed using the Phred basecaller (Ewing et al. (1998) Genome Res. 8:175-185, assembled with Phrap assembler and contig assemblies then viewed with Consed (at http://www.phrap.org/).

EXAMPLE 2

Characterization of Arabidopsis Ftn2 Gene and Protein

This example describes the identification, isolation, and characterization of an *Ftn2* gene from Arabidopsis.

Identification of Arabidopsis *arc6* mutation

Available mapping data for the *arc6-1* mutant (Marrison et al. (1999) Plant J. 18:651-662; Rutherford (1996) In Dept of Biology, University of York, York 161-209) suggested that the mutation is located on chromosome 5, between the markers m247 and DFR, very close to the marker g4028. The tblastn search of Arabidopsis genome with the Synechococcus sp. WH8102 Ftn2 cell division gene (as described below, and in Koksharova and Wolk (2002) J Bacterial: in press) in preparation) (see Table 3 below) revealed a homologue on chromosome

5, At5g42480 (Accession number NM_123613) in close proximity to the genetic markers mentioned above. This gene was designated *AtFtn2*. The gene was sequenced from the wild-type and *arc6-1* plants, where the sequence included the flanking regions of about 500 nt 5' and 200 nt 3'. Compared to the wild type *AtFtn2* gene, *arc6* showed two nucleotide differences. The first difference was found at position 1141: T in arc6, C in the WT-Ws, close to the end of exon 3, resulting in a premature stop codon (TGA) in *arc6* and a truncated protein of 324 amino acids (Figs. 1, 2). The second difference was found at position 1790: G in arc6, A in WT-Ws. This difference was attributed to slightly different genetic backgrounds of arc6-1 (Ws-2) and the WT used (Ws, unknown subtype), since the published sequence of WT-Columbia (NM_123613) was identical to that of *arc6* in this area.

Sequencing of *arc6-2* and *arc6-3* revealed a mutation identical to that in *arc6-1*. To further confirm this result and to ascertain that the *arc6-2* and *arc6-3* were not accidentally mislabeled or confused with *arc6-1*, the region of interest was sequenced from additional *arc6-2* and *arc6-3* mutants obtained from the Nottingham Arabidopsis Stock Centre (seed stock number N286 and N287, respectively). These mutants, too, carried the same mutation as *arc6-1*.

The *arc6* mutation is rescued by a wild-type copy of *AtFtn2*

Genomic *AtFtn2* DNA, containing about 0.5kb 5' and 0.2 kb 3' region, was introduced into the *arc6-1* and WT plants via *Agrobacterium*-mediated floral-dip transformation. T1 plants carrying the selection marker were assessed for leaf chloroplast size and numbers. Most T1 plants of the *arc6-1* background showed less severe plastid phenotypes than the parent *arc6-1* mutant. Plastids were more numerous and smaller, and approximately 80% of the T1 plants had WT-like phenotypes (Table 1). A majority of the plants with the WT background had normal (WT-like) phenotypes, even though some plants showed occasional clusters of cells with enlarged, irregularly shaped chloroplasts.

Table 1
Leaf mesophyll chloroplast phenotypes in T1 plants carrying AtFtn2 transgene.

Genetic background	# plants total	WT-like phenotypes	Intermediate plastid size,	Severe chloroplast

			number	phenotype
WT Ws	205	191	0	14
Arc6-1	120	97	18	5

Characterization of *AtFtn2* gene and protein:

a plastid-targeted protein with an unconventional DnaJ-like domain

The *AtFtn2* genomic sequence has 6 exons (Figure 3). The presence of EST and full length cDNA in the sequence database (Table 3 below) indicates that the gene is expressed. Both the predicted and the experimentally determined full length cDNA coding sequences (Table 3 below) have 2406 nt encoding a protein of 801 aa, with putative N-terminal chloroplast targeting sequence of 67 aa predicted by TargetP. Chloroplast targeting was also predicted by Predotar (targeting scores 0.738 and 0.979 for TargetP and Predotar, respectively).

A search for protein motifs with InterProScan revealed a putative DnaJ domain (AtFtn2 residues 89-153), InterPro accession IPR001623, Pfam conserved domain pfam00226. However, ClustalW alignment of this domain with all predicted DnaJ domains from the Pfam database (277 sequences) revealed that the central Histidine-Proline-Aspartate (HPD) motif typical for DnaJ proteins is not present in *AtFtn2* or in other plant and cyanobacterial *Ftn2* homologues (Figure 4). In addition to the DnaJ-like domain, the Pfam-HMM search identified a putative myb domain (residues 677-690, see Figure 4) albeit with low expectation value (0.63). Sequence alignment with myb domains from the Prosite database indicated that only a second half of the putative myb domain is present in AtFtn2.

Annotation for AtFtn2 in the MIPS database (mips.gsf.de/cgi-bin/proj/thal/gv_report?mdh9+At5g42480) stated that AtFtn2 is a membrane protein Furthermore, preliminary results from the ongoing proteomics project at Michigan State University, which is directed at identifying components of the chloroplast envelope, indicated that AtFtn2 is present in the envelope membrane fraction from isolated *Arabidopsis* chloroplasts. Up to three putative transmembrane helices were predicted, using different software tools (Table 2).

Table 2
Putative transmembrane (TM) regions in AtFtn2

Prediction program	TM region
HMMTOP	297 - 314, 615 - 632
DAS	207 - 215, 354 - 356, 621 - 630
TopPred 2	56 - 76, 295 - 315, 615 - 635
Tmpred	46 - 71, 297 - 313, 619 - 634
SOSUI	615 - 636
Split	615 - 634
TMHMM	None

Plastid-dividing cytoskeletal FtsZ rings and filaments are severely disrupted in *arc6*

Immunoblots showed that levels of the cytoskeletal, chloroplast-dividing proteins AtFtsZ1 and AtFtsZ2 were slightly lower in *arc6-1* and *arc6-2* mutants compared to the WT. Immunofluorescence labeling of *arc6* leaf chloroplasts was done with antibodies specific to AtFtsZ1 and AtFtsZ2. The mmunolabeling was highly specific for the target proteins, as indicated by the controls where the antibodies were omitted, as well as by previous results (Vitha et al. (2001) J Cell Biol. 153:111-119). These earlier results also demonstrated that AtFtsZ1 and AtFtsZ2 proteins are colocalized in FtsZ filaments and rings, in both the current set of WT and mutant plants (McAndrew et al. (2001) *Plant Physiol.* 127:1656-1666; (Vitha et al. (2001) J Cell Biol. 153:111-119).

In WT leaf chloroplasts, AtFtsZ1 and AtFtsZ2 are localized in rings at mid-plastids. In contrast, *arc6* plastids show numerous short and disorganized AtFtsZ filaments. To investigate the possibility that the fragmentation and disruption of FtsZ rings and filaments is a consequence of the gross enlargement of the chloroplast rather than being directly related to the *arc6* mutation, AtFtsZ localization patterns were analyzed in several mutant or transgenic plants with very large chloroplasts. Plants carrying antisense or overexpression constructs of *AtFtsZ1-1*, *AtFtsZ2-1* or *AtMinD*, the chloroplast division-site determining factor (Colletti et al. (2000) Curr Biol. 10:507-516), as well as the *arc3* mutant of Arabidopsis (Marrison et al. (1999) Plant J. 18:651-662) were used. The results indicate that intact FtsZ rings and/or long FtsZ1 and FtsZ2 filaments can assemble in large chloroplasts as well as in the WT. However, overexpression of AtMinD caused disruption and fragmentation of FtsZ rings and filaments,

an effect somewhat similar to the FtsZ pattern in *arc6*. This is consistent with the suggested role of AtMinD in preventing FtsZ ring assembly at improper sites (Dinkins et al. (2001) Planta. 214:180-188; Kanamaru et al. (2000) Plant Cell Physiol. 41:1119-1128).

EXAMPLE 3

Ftn2 Homologues in Other Plants and Cyanobacteria

This example describes the identification of other Ftn2 homologues in other plants and in cyanobacteria.

Tblastn search with AtFtn2 and *Synechococcus* sp. WH8102 Ftn2 proteins as a query revealed homologues in all publicly available fully sequenced cyanobacterial genomes and also in rice (*Oryza sativa*) non-annotated genomic DNA sequence (Table 3). Additionally, a number of ESTs representing *ftn2* homologues from vascular plants, as well as a moss (*Physcomitrella patens*) and a fern (*Ceratopteris richardii*) homologue, were identified (Table 3). No *ftn2* homologues were found in non-cyanobacterial prokaryotes.

Table 3: Homologues of Ftn2

Results of tblastn search with the *Arabidospis* AtFtn2 protein sequence. For ESTs, the reading frame and the area of match with AtFtn2 are indicated.

Species	ORF/Gene name	Accession # (DNA)	Protein Accession #	Type[2]	Frame, tblastn match with Arabidopsis ARC6
Arabidopsis thaliana	At5g42480[1] *ARC6*	NM_123613 AB016888 [13]	NP_199063 BAB10489	Gen	
Arabidopsis thaliana		AI998415		EST	-3; 642-801
Arabidopsis thaliana	At5g42480	AY091075	AAM13895	cDNA	Full length cDNA
Medicago truncatula		AL382914		EST	+3; 623-717
Medicago truncatula		AL382915		EST	+3; 693-801
Medicago truncatula		BI268376		EST	+3; 33-239
Medicago truncatula		AW696905		EST	+2; 95-121 +3; 121-258 +1; 244-277
Gossypium arboreum		BQ410207		EST	-2; 679-798
Gossypium arboreum		BQ410206		EST	+2; 679-801

Glycine max		AW472683		EST	+2; 173-221
Solanum tuberosum		BE472035		EST	+3; 1-177
Beta vulgaris		BQ490457		EST	+3; 585-691
Populus balsamifera		BI120337		EST	+1; 316-409
Mesembryanthemum crystallinum		AI043508		EST	+1; 747-801
Oryza sativa		AU095068		EST	+3; 501-576
Oryza sativa		AU183658		EST	+3; 286-381
Oryza sativa		AU058418		EST	+3; 286-384
Oryza sativa [7]		BK000999		cDNA	
Triticum aestivum		BQ238871		EST	+3; 710-801
Triticum aestivum		BJ263824		EST	-3; 679-801
Triticum aestivum		BJ258222		EST	+1; 129-287
Triticum aestivum		BE490117		EST	+3; 186-362
Triticum monococcum		BQ169059		EST	-2; 708-801
Triticum monococcum		BG607272		EST	+1; 267-413
Hordeum vulgare		BJ482132		EST	+2; 165-294
Hordeum vulgare		AJ463103		EST	+2; 708-801
Hordeum vulgare		AJ485539		EST	+1; 666-784
Hordeum vulgare		BJ464825		EST	+2; 249-457
Hordeum vulgare		AJ485537		EST	+1; 666-801
Hordeum vulgare		BI949952		EST	+3; 666-801
Hordeum vulgare		AV833644		EST	+3 ; 290-472
Hordeum vulgare		AV921157		EST	-3; 683-801
Sorghum bicolor		BE917942		EST	+1; 671-801
Sorghum bicolor		BE918523		EST	+2; 613-752
Zea mays		BQ048486		EST	-1; 200-366
Zea mays		BM498278		EST	+3; 34-185
Zea mays		BM498757		EST	-3; 211-358
Zea mays		AW331058		EST	+2; 673-798
Ceratopteris richardii		BE641509		EST	+3; 305-488
Physcomitrella patens		BI437111		EST	+2; 669-799
Protochlorococcus marinus MED4	Contig1, Gene_533 [5]			Gen	
Protochlorococcus marinus MT9313	Contig1, gene2677 [6]			Gen	
Synechococcus sp. PCC 7002	Contig05130 2-306[3]			Gen	
Synechococcus sp. PCC 7942	*Ftn2*	AF421196	AAL16071	Gen	
Anabena PCC 7120	all2707	AP003590[8] NC_003272[9]	BAB74406 NP_486747	Gen	
Nostoc punctiforme ATCC 29133	Contig493 Gene 84 [4]			Gen	

Synechocystis sp. PCC 6803	sll0169	NC_000911[10] D63999[11]	NP_441990 BAA10060	Gen	
Arabidopsis thaliana	At3g19180	AY074283	AAL66980	cDNA	Full length cDNA
Arabidopsis thaliana	At3g19180	NC_003074[12]	NP_188549	Gen	
Synechococcus sp. WH8102	Gene 3082				
Thermosynechococcus elongatus BP-1	tlr0758			GEN	
Trichodesmium erythraeum IMS101	Contig97 Gene 8639			GEN	
Chlamydomonas reinhardtii	genie.294.6 (Scaffold294, nt 47288-51078)			GEN	
Prunus persica (peach)		BU046755		EST	+1; 315-508
Helianthus annuus		BU035730		EST	+1; 627-801
Helianthus annuus		BQ977057		EST	+1; 664-801
Populus tremula		BU889000		EST	+1; 613-759

[1] Standard Arabidopsis ORF name (http://arabidopsis.org/info/guidelines.html)
[2] Type of DNA sequence: EST (Expressed Sequence Tag), cDNA (full length cDNA), Gen (Genomic DNA)
[3] Unfinished fragment of the genome, Joint Genome Institute (JGI)
[4] Draft analysis; http://genome.ornl.gov/microbial/npun/31may01/npun.html
[5] draft analysis http://genome.ornl.gov/microbial/pmar_med/
[6] Draft analysis http://genome.ornl.gov/microbial/pmar_mit/
[7] AAAA0100502 Predicted Gen sequence from shotgun sequencing data, see Methods; BK000999 cDNA sequence
[8] complement (211130..213526)
[9] complement (3300430..3302826)
[10] complement (2314780..2316924)
[11] complement (47521..49665)
[12] bases 6632806..6639031
[13] bases 64077..67114; gene id: MDH9.18

In order to obtain putative protein sequence of the rice Ftn2 from the genomic sequence, results from several gene prediction programs, EST database records and tblastn alignment with AtFtn2 (see Example 1) were combined. It is contemplated that the rice Ftn2 (OsFtn2) is encoded on the reverse strand of the contig (Accession AAAA01000502) and has

7 exons (8785-8486, 8104-7874, 7743-7546, 7380-7120, 7022-6158, 5923-5790, 5510-5217). The predicted protein has 760 amino acids.

TargetP analysis of the full length rice and partial potato Ftn2 sequences, for which the N-terminal portions were complete and included the initial M, identified putative chloroplast targeting signals of 40 and 76 aa, respectively, with prediction scores 0.961 and 0.583. Predotar predicted chloroplast targeting for the rice (score 0.928) but not potato Ftn2 (score 0.032).

ClustalW alignment of full and partial Ftn2 protein sequences (Figure 5) showed that the N- terminal, and to a lesser degree also the C-terminal, regions of these proteins are conserved and separated by a highly divergent central area (Figure 3B). The cyanobacterial homologues shared approximately 20% identity and 40 % similarity with AtFtn2, while scores for the rice homologue were 47% and 68%, respectively (Table 4).

Table 4

Similarity and identity scores of Ftn2 homologues compared to *Arabidopsis* AtFtn2.
Sequence alignment does not include the N-terminal portion with chloroplast targeting signals - the first 74 amino acids of AtFtn2 were removed

Species	% Identities	% Similarities
Anabena PCC 7120	19	38
Nostoc punctiforme ATCC 29133	19	39
Protochlorococcus marinus MED4	15	38
Protochlorococcus marinus MT9313	16	40
Synechocystis sp. PCC 6803	19	40
Synechococcus WH8102	17	38
Oryza sativa	47	68

Tblastn search with AtFtn2 also revealed an *Arabidopsis* membrane protein of unknown function, At3g19180 (Table 3), which showed a 21% and 44% identity and similarity, respectively, with AtFtn2. This protein is 970 aa long and contains an N-terminal targeting sequence. However, the targeting prediction is controversial: it is either a chloroplast (TargetP score 0.723) or a mitochondrial (Predotar score 0.846) target. A number of ESTs from maize, barley, sorghum, wheat and tomato were found in tblastn search using At3g19180 as a query.

EXAMPLE 4

Materials and Methods Utilized to Identify and Characterize Cyanobacterial *Ftn2* genes

This example describes the materials and methods used to identify and characterize cyanobacterial *Ftn2* genes. The designation "*Ftn2*" refers to the mutant phenotype in which cell division is inhibited, resulting cells that are longer than wild-type cells, or filamentous in appearance. In classical studies of *f*ilamentous *t*emperature-*s*ensitive mutants of *E. coli* affected in cell division (Bramhill D (1997) Annu. Rev. Cell. Dev. Biol. 13:395-424), the corresponding genes were designated *fts*; therefore, by analogy, the cell division mutants isolated as described below were initially designated FTN-mutants (Filamentous, TransposoN-derived), and the corresponding genes, *Ftn*.

Bacterial strains, plasmids, and culture conditions

Wild type *Synechococcus* sp. strain PCC 7942 and its derivatives (Table 5) were grown in BG11 medium (Rippka RJ, et al. (1979) J. Gen. Microbiol. 111:1-61). Wild type *Anabaena* sp. strain PCC 7120 and its derivatives were grown in media with or without nitrate supplementation as described by Hu et al. (Hu NT et al. (1982) Virology 114:236-246). Derivative strains were grown in the presence of appropriate antibiotics. Cyanobacterial cells were grown in 125-ml Erlenmeyer flasks at 30 °C in the light (about 3,500 ergs cm^{-2} s^{-1}) on a rotary shaker. Growth and plasmid transformation of *E. coli*, selection, and testing of transformants were performed as described (Sambrook J et al. (1989) Molecular Cloning, a laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). Plasmids with or without transposon Tn*5*-692 were transferred to PCC 7942 and to *Anabaena* sp. strain PCC 7120 by conjugation with *E. coli* strain HB101 bearing pRL443, pRL528, and pRL692 (Cohen MF et al. (1998) Methods Enzymol 297:3-17). Plasmids pRL2462 and pRL2463 (see Table 5) were introduced into *Synechococcus* sp. strain PCC 7942 by transformation (Koksharova O et al. (1998) Plant Mol. Biol. 36:183-194).

Table 5

Cyanobacterial strains and plasmids used

Strain or plasmid	Derivation and/or relevant characteristics	Source
Synechococcus sp. strain		
PCC 7942	Wild type	L. Sherman
FTN2	$Sm^rSp^rEm^r$; Tn*5*-692 mutant	This study
FTN6	$Sm^rSp^rEm^r$; Tn*5*-692 mutant	This study
Anabaena sp. strain		
PCC 7120	Wild type	R. Haselkorn
$FTN2_A$	Nm^r; PCC 7120::pRL2471	This study
$FTN6_A$	Nm^r; PCC 7120::pRL2474	This study
Plasmids		
pRL443	Ap^rTc^r; Km^s derivative of RP4	(19)
pRL498	Km^r; positive selection cloning vector	(20)
pRL528	Cm^r; bears *avaIM* and *eco47IIM*	(19)
pRL692	$Em^rSm^rSp^r$, bears Tn*5*-692	This study
pRL2462	Sm^rSp^r; chromosomal DNA from FTN2 cut with *Sal*I, religated, and transformed to *E. coli*	This study
pRL2463	Sm^rSp^r; chromosomal DNA from FTN6 cut with *Sal*I, religated, and transformed to *E. coli*	This study
pRL2464	Ap^r; pBluescript®SK(+)(Stratagene) cut with *Xba*I and ligated to *Spe*I-*Spe*I fragment from pRL2463	This study

pRL2465	Ap^r; pBluescript®SK(+) cut with *Xba*I and *Sal*I, ligated to *Xba*I-*Sal*I fragment from pRL2463	This study
pRL2466	Ap^r; pBluescript®SK(+) cut with *Xba*I and *Sal*I, ligated to *Xba*I-*Sal*I fragment from pRL2462	This study
pRL2468	Ap^r; pBluescript®SK(+) cut with *Spe*I and *Sal*I, ligated to *Spe*I-*Sal*I fragment from pRL2462	This study
pRL2471	Km^r; pRL498 with truncated PCR copy of $Ftn2_A$	This study
pRL2474	Km^r; pRL498 with truncated PCR copy of $Ftn6_A$	This study
PRL2733	Sm^rSp^r; chromosomal DNA of FTN2 cut with *Bln*I, religated and transformed to *E. coli*	This study

[a] Ap, ampicillin; Em, erythtomycin; Km, kanamycin; r, resistant; s, sensitive; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline.

Transposon mutagenesis of *Synechococcus* sp. strain PCC 7942

Transposon Tn*5*-692 (in plasmid pRL692: GenBank accession no. AF424805) is a derivative of transposon Tn*5* that confers resistance to erythromycin (Em), spectinomycin (Sp), and streptomycin (Sm); contains a pMB1 *oriV*; and bears mutations (Zhou M et al. (1998) J Mol. Biol. 276:913-925) that increase its rate of transposition ca. 100-fold relative to pRL1058 (Wolk CP et al. (2000) Heterocyst formation in Anabaena, pp. 83-104 In: Y.V. Brun and L.J. Shimkets (ed), Prokaryotic Development, American Society for Microbiology, Washington). Plates with filter-borne cells were incubated 48 h at 30 °C (light intensity, 1,500 ergs cm^{-1} s^{-1}), and the filters then transferred onto solid BG11 medium containing 10 μg ml^{-1}, each, of erythromycin and spectinomycin. Antibiotic-resistant colonies appeared 10-15 days later.

Mutant selection and microscopy

Mutants exhibiting a filamentous phenotype spread extensively on solid medium. Mutant cells grown in liquid medium were examined by microscopy, and photographed at 400 and 800 times magnification with a Zeiss (Carl Zeiss, D-7082, Oberkochen, Germany) Axiophot microscope. Samples were prepared for electron microscopy and micrographed by S. Burns, MSU Center for Electron Optics.

Cloning and sequencing of *Synechococcus* PCC 7942 *Ftn* genes

Transposon Tn*5*-692 contains an *oriV* active in *E. coli*. Therefore, to clone PCC 7942 DNA contiguous with the transposon, DNA recovered from FTN2 was cut separately with *Sal*I and *Bln*I, whose targets are absent from the transposon, circularized with T4 DNA ligase, and transformed to *E. coli* DH5α, yielding plasmids pRL2462 and pRL2733, respectively, and DNA recovered from FTN6 was cut with *Sal*I, circularized, and transformed to DH5α, yielding pRL2463. Fragments contiguous with the transposon were subcloned to pBluescript SK(+) (Stratagene, La Jolla, California 92037, USA) and sequenced. To compare sequences of *Ftn2* and *Ftn6* from the FTN mutants and from wild-type *Synechococcus* sp. strain PCC 7942, genomic DNA from wild-type PCC 7942 was isolated as described by Koksharova et al. (Koksharova O et al. Plant Mol. Biol. 36: 183-194) and PCR amplifications and sequencing were performed with gene specific primers (Table 6). With the exception of the final 183 bp of *Ftn2*, which were sequenced only from pRL2733 as template, all portions of *Ftn2* and *Ftn6* were sequenced on both strands of DNA derived from a transposon recovery and on both strands of DNA PCR-amplified from wild type DNA; where there was any possible inconsistency, multiple independently PCR-amplified fragments of DNA were sequenced. The sequences of *Ftn2* and *Ftn6* have been submitted to GenBank under accession nos. AF21196 and AF21197, respectively.

TABLE 6

DNA primers for PCR and sequencing of *Ftn2* and *Ftn6* of *Synechococcus* sp. PCC 7942

Primers	Used for PCR	Used for sequencing
Ftn2-specific		
Cpw267 5'- CCGAATTCTCTGTGTTGGCG-3' (D)	+	+
Cpw268 5'-AAGCTTCGTACAGACCCTGCTGAC-3' (R)	+	
Cpw338 5'-GGTAAGTTGACGGTCAAG-3' (D)	+	+
Cpw339 5'-CGATAGGGCCGTAGCTGTC-3' (R)	+	+
Cpw355 5'-GGTTAACTTGTGATCGAAC-3' (R)	+	+
Cpw376 5'-GCAGCCAGTCTGCCCTAG-3' (D)		+
Cpw377 5'-GCGCAGTCCTTTCTTGAGG-3' (R)		+
Cpw384 5'- CTGACCGGTGAGGTTCTGC-3' (D)		+
Cpw386 5'- CCAGGAATCGCTGAACATTC-3'(R)		+
Cpw387 5'-GCGATCGCGGTAGCTTTCGG-3' (R)		+
Cpw400 5'-CTAGGCAGTGTACGTTC-3' (D)		+
Ftn6-specific		
Cpw269 5'-CCGAATTCGTGACCTCTACCCGTACTGC-3'(D)	+	+
Cpw270 5'-CCAAGCTTCGTTTTATAAAGGCGCTCAG-3'(R)	+	+
Cpw340 5'-CTGCTCGTGAGCAATTTGC-3' (D)	+	+
Cpw341 5'-CCGTTCTGAAAGGCTC-3' (R)	+	+
Cpw396 5'-CAGTGAATTGTAATAC-3' (D)		+
Cpw398 5'-GAAATAGCCATCGCGAGC-3'(R)		+

Insertional inactivation of *Ftn2* and *Ftn6* orthologs in *Anabaena* sp. strain PCC 7120

Orthologs $Ftn2_A$ of *Ftn2* and $Ftn6_A$ of *Ftn6* were identified in the genome of *Anabaena* sp. strain PCC 7120 by tblastn and blastn search against the complete Anabena genome database at the Kazusa DNA Research Institute (kazusa.or.jp/cyano/anabaena). Copies of (i) $Ftn2_A$ and (ii) $Ftn6_A$ truncated at both ends were prepared by PCR with isolated genomic DNA of PCC 7120 as template using:

(i) CPW263, 5'-CCGAATTCGTGGCAGTGGAAAATCGTGGG-3', as direct primer and CPW264, 5'-CCGAATTCCACTTGCACGATTGGGATC-3', as reverse primer and;

(ii) CPW265, 5'-CCGAATTCGCCCTACTCATTAACTATAG-3', as direct primer and CPW266, 5'-CCGAATTCCGGAGCGATCGCTTGTTTG-3', as reverse primer.

The PCR-generated copies were cloned in the *Eco*RI site of pRL498 (16), and the clones transferred by conjugation to wild-type PCC 7120, with selection on AA + nitrate agar medium (Fink A (1999) Physiological Rev. 79:6025-6032) containing 25 μg neomycin ml^{-1}.

Southern hybridization

Southern hybridization was performed as described by Sambrook et al. (45), with digoxigenin-dUTP-labelled probes (DIG DNA Labeling Kit, Roche Diagnostics Corp., Indianapolis, IN). Probes for Southern analysis were prepared by PCR with the following primers: *Ftn2*, CPW 267 and CPW 268; *Ftn6*, CPW 269 and CPW 270 (see Table 2); $Ftn2_A$, CPW263 and CPW264; and $Ftn6_A$, CPW265 and CPW266 (see above).

EXAMPLE 5

Identification, Isolation, and Characterization of Cyanobacterial *Ftn2* Gene and Protein

This example describes the identification, isolation, and characterization of an *Ftn2* gene from cyanobacteria.

Transposon mutagenesis and analysis of *Ftn* genes of *Synechococcus* sp. strain PCC 7942

When *Synechococcus* sp. strain PCC 7942 was mutagenized with transposon Tn*5*-692, about 3000 Em^rSp^r, dense, round mutant colonies with regular margins were accompanied by

39 spreading colonies with irregular borders that were comprised of very elongated cells. In classical studies of *f*ilamentous *t*emperature-*s*ensitive mutants of *E. coli* affected in cell division (6), the corresponding genes were designated *fts*. Therefore, by analogy, the transposon-derived cell division mutants were designated FTN-mutants (Filamentous, TransposoN-derived) and the corresponding genes, *Ftn*. Two such mutants whose irregular colonies are composed of cells that are longer than wild-type cells, designated FTN2 and FTN6, were further characterized. The cells of FTN2 are very long, up to 100-fold the length of wild-type cells, whereas the cells of FTN6 are only up to 20 times longer than those of the parental strain. Because the septation of these serpentine cells was not easily visualized by light microscopy, the cells were negatively stained with uranyl acetate, and examined by electron microscopy. The cells of both mutants usually divided asymmetrically. Plasmids pRL2462, pRL2463, and pRL2733 contain transposon DNA and contiguous PCC 7942 DNA. The first two were transformed to PCC 7942. All spectinomycin- and erythromycin-resistant transformants were filamentous, establishing that the mutations were closely linked to the transposon. Mutants FTN2 and FTN6 are completely segregated.

DNA contiguous with the transposon was subcloned from pRL2462 to pBluescript SK(+) as *Xba*I-*Sal*I and *Spe*I-*Sal*I fragments, producing plasmids pRL2466 and pRL2468, respectively, and from pRL2463 to pBluescript SK(+) as *Xba*I-*Sal*I and *Spe*I-*Spe*I fragments, producing plasmids pRL2465 and pRL2464, respectively. Part of plasmid pRL2733 was sequenced with primers. The expected 9-bp duplication adjacent to the site of insertion of the transposon was found in the case of FTN6, but the same two transposon-proximal 9-bp sequences differed at one position (TGCAGGCG[C/T]) as recovered from FTN2. To resolve this difference, and to determine whether the sequences determined with the transposon-mutated genes were identical to the wild-type sequences, both genes were amplified piecewise by PCR from wild-type PCC 7942 and the products of PCR were sequenced. Independent PCR amplifications confirmed that the sequence TGCAGGCGC is adjacent to the position of the transposon in *Ftn2*.

In FTN2 and FTN6, the transposon was inserted in single-copy open reading frames (ORFs) that were denoted *Ftn2* and *Ftn6*. *Ftn2* predicts a 631-amino acid protein (see Figure 6, panel B) that shows greatest similarity to the predicted products of an ORF designated $Ftn2_A$

from *Anabaena* sp. strain PCC 7120 (bp 3302826-3300430 in the chromosome (see Figure 8); BLAST score, 278; Expect = 3 x 10^{-75}; [1]), a *Nostoc punctiforme* ORF (BLAST score, 263; Expect = 1 x 10^{-70}), and presumptive gene *sll0169* of *Synechocystis* sp. strain PCC 6803 (BLAST score, 218; Expect = 2 x 10^{-55}).

The InterProScan program (http://www.ebi.ac.uk/interpro/scan.html) shows the presence in Ftn2 of a DnaJ N-terminal domain (amino acid residues 6-70) and a single TPR repeat (amino acid residues 136-169. The Prosite-Protein against PROSITE program (http://ca.expasy.org/tools/scnpsite.html/) shows the presence in Ftn2 of a leucine zipper pattern (amino acid residues 234-255; Table 7). Ftn2 and its cyanobacterial and plant orthologs show the presence of a DnaJ N-terminal domain, but are otherwise, as are Ftn6 and its orthogs, dissimilar from the products of known division-related genes (Bramhill D (1997) Annu. Rev. Cell. Dev. Biol. 13:395-424).

Table 7
Characteristics of Ftn2 and its homologs

Protein and organism	Number of aa	MW (kDa)	pI	Domains or pattern
Ftn2				
Synechococcus sp. PCC 7942	648	72.4	5	1. DnaJ N-terminal domain (aa 6-70) 2. TPR repeat (aa 136-169) 3. Leucine zipper (aa 234-255)
$Ftn2_A$				
Anabaena sp. PCC 7120	798	90.1	6.3	1. DnaJ N-terminal domain (aa 16-80)
Ftn2 ortholog				
Nostoc punctiforme	768	87.4	6.8	1. DnaJ N-terminal domain (aa 16-80) 2. ATP/GTP binding site motif A (P-loop) (aa 566-573)

S110169				
Synechocystis	714	79.4	4.7	1. DnaJ N-terminal domain (aa 6-70)
PCC 6803				
AB016888				
Arabidopsis	801	88.3	4.6	1. DnaJ domain profile (aa 89-153)
thaliana				2. Myb DNA-binding domain (aa 677-690)

aa = amino acid residues

The gene *Ftn6* predicts a 152-amino acid protein that shows greatest similarity to an ORF from contig 630 of *N. punctiforme* (BLAST score, 80; E = 3 x 10^{-16}), an ORF from *Anabaena* sp. strain PCC 7120 denoted *Ftn6*$_A$ (bp 1903579-1902896 in the chromosome; BLAST score, 77.8; E = 10^{-15}) and a predicted protein, Sll1939, from *Synechocystis* sp. strain PCC 6803 (BLAST score, 59; E = 1 x 10^{-08}).

Inactivation of the *Ftn*$_A$ genes of *Anabaena* sp. strain PCC 7120

Anabaena sp. strain PCC 7120, a filamentous cyanobacterium, is capable of cellular differentiation ((Wolk CP et al. (2000) Heterocyst formation in Anabaena, pp. 83-104 In: Y.V. Brun and L.J. Shimkets (ed), Prokaryotic Development, American Society for Microbiology, Washington). Experiments to mutate the *Anabaena* sp. orthologs *Ftn2*$_A$ and *Ftn6*$_A$ were undertaken to observe whether the effects of inactivating these genes would be similar to those observed in *Synechococcus*, and whether there might be an effect on differentiation.

A truncated, PCR-generated copy of each gene was cloned in pRL498, producing plasmids pRL2471 and pRL2474, respectively. Cells of *Ftn2*$_A$ and *Ftn6*$_A$ *Anabaena* sp., i.e., of PCC 7120::pRL2471 and PCC 7120::pRL2474, grown in the presence of nitrate were often up to twice as long as cells of the wild-type strain. In medium free of combined nitrogen, both mutants formed very elongated vegetative cells (those of *Ftn2*$_A$ were up to 60-fold longer than those of the wild-type strain); heterocysts of nearly normal size (but also sometimes up to 4-fold

larger, with an increase in both length and width); and also enlarged akinete-like cells. Because mutant $FTN2_A$ is not completely segregated, gene $Ftn2_A$ may be important for viability of *Anabaena*. Mutant $FTN6_A$ is completely segregated.

EXAMPLE 6

Identification of ARC5

This Example describes the identification of the Arabidopsis *ARC5* gene.

The *arc5* mutation was induced by EMS mutation in Arabidopsis strain Landsberg *erecta* and identified as a chloroplast division mutant by microscopic screening (Robertson et al., (1996) Plant Physiol **112**(1): 149-59. Phenotypes were analyzed as previously described (Osteryoung, K. W. *et al.* (1998) *Plant Cell* 10, 1991-2004), except that the images were recorded with a Coolpix 995 digital camera (Nikon Corporation, Tokyo, Japan). *arc5* cells were found to have about 5 to 10 chloroplasts per cell. The chloroplasts are larger than in wild type. Constricted chloroplasts were frequently found. The proportion of constricted chloroplasts varied in different plants.

The *arc5* mutation was previously mapped between markers nga 162 (20.6 cM) and AtDMC1 (32.6 cM) on chromosome 3 (Marrison et al., 1999 Plant J **18**(6): 651-62). To fine-map the position of *arc5*, an F_2 population was generated from a cross between *arc5* and Col-0 wild type. 1720 mutant plants out of 7000 F_2 plants were selected and their DNA was extracted for PCR marker-based mapping. Markers were generated using the primer sets shown in Table 8:

Table 8

Primer Sequences

BAC Clone name	Primer sequences for PCR	Marker type
MDC8	GATTAATGAGACTATATATGAGAG and ATCTGCATAACTTCAATTGAACTG	INDEL
MCB22	GAACCCCCAGAATATCAACATC and GCTCTGATGGTGATTCTGGTAAC	INDEL
MVI11	GTAGCATTCTTTAGAGATTGATCTAG and	INDEL

	TATTCGAGTTTGAAATTATGATTTATGC	
MLD14	GCTACAGTTCTCAACCGGTAAATC and CATAAGCTTTTATGCTCCAAAATAGTCTC	INDEL
T31J18	CTTGATCTTGTGTTCTGACATCTC and CTAAACTATTCACAAATGCCATAGACG	CAPS, cut by DraI
MMB12	AGCCGTCTTGTCCCATCATTAAAG and GCACAAACAAACAGGGTCAATAGTTA	CAPS marker, cut by EcoRV
F16J14	TTAAAGTGAAGCTTAAGCAGAGG and CATTGTTAGAAAGTCAACACTTTG	INDEL
MSA6	GCAAGACATAACCAATGAACAAG and GACACGTATGCGTTTCTAAGAG	INDEL
MAL21	CTCCAACTTCAAGCAAAACGGATG and CTCTGTTTTTTGGGCTAGTGATGG	INDEL
MPN9	GCATACCCAATATCCTTTGTGC and GATAGTATAACCAGAGGTTGGAG	CAPS marker, cut by Tsp509I

The results indicated that *arc5* was located either on BAC clone MMB12 or MPN9, which overlap. The following three additional markers were generated, but no recombination between these and *arc5* was observed.

Table 9

Primer Sequences

BAC Clone name	Primer sequences for PCR	Marker type
MMB12	GAATCTTCTCAAACTGAAATCCACC and TCGAAAGGAAGATCGGTGAACC	CAPS marker, cut by TaqI
MPN9	GATTGTGCTATGGTTCAGGAGTTC and CATCAGCTATAACCTCCTCAGTG	CAPS marker, cut by AccI
MPN9	ACTGACTATAAGGACCCCTCAAAC and GTTGACCATAATTCATCCACCACTATTA	INDEL but cut by HindIII

The mapping studies narrowed down the interval of chromosome III containing *arc5* to a 92-kb region comprising DNA spanning the overlap between MMB12 and MPN9.

To identify the DNA corresponding to *arc5*, BAC insert DNA from MMB12 and MPN9 was double-digested with HincII and HindIII. The digested fragments were inserted between 35S promoter and OCS terminator in the plant transformation vector pART27 (Gleave, 1992 Plant Molecular Biology 20: 1203-1207) to make a small transformable antisense/sense library. The library was transferred to *Agrobacterium tumefaciens* strain GV3101, and used to transform wild type Arabidopsis plants (Col-0) by floral dipping. 120 transformants were screened by microscopy for chloroplast division defects. Two plants were found to have only a few large chloroplasts per cell. The fragments between the 35S promoter and OCS terminator in the transgenes from these two plants were amplified by PCR and sequenced. One plant carried a transgene containing a fragment of the BAC backbone DNA, and another fragment from At3g19730 in the antisense orientation. The other plant also carried the same fragment from At3g19730 in the antisense orientation, as well as a second fragment from At3g19760. Based on these findings, it was predicted that the *arc5* gene corresponded to At3g19730, which is predicted to be a dynamin-like protein. To confirm the plastid division phenotype in the transgenic plants was from this gene, an antisense transgene was constructed containing the fragment from At3g19730 carried by the two plants described above, and transformed into wild-type Arabidopsis (Col-0). 80 transformed plants were screened under the microscope. 20% of the transformants displayed fully expanded cells with fewer and larger chloroplasts than in wild type. These phenotypes resembled those in *arc5*. This further confirmed that At3g19730 functioned in chloroplast division and is *ARC5*.

In the NCBI database, At3g19720 and At3g19730 were annotated as a single gene, MMB12.21. Based on the alignment of MMB12.21 to the other dynamin-like proteins in Arabidopsis, it appeared that NCBI's annotation of this region was more accurate. Thus, they may be referred to as At3g19730/At3g19720; moreover, the annotated start codon for At3g19730 and stop codon for At3g19720 represent the true start and stop codons of this gene. The whole region of MMB12.21 in the *arc5* mutant, and well as in wild-type Landsberg *erecta,* was sequenced. The data revealed a G-to-A mutation (C-to-T on the

opposite strand) at nucleotide 60730 of MMB12. This mutation caused a change from the tryptophan codon "TGG" to the stop codon "TAG", in the 5th exon of MMB12.21. This mutation also created a new restriction enzyme cutting site —Xba I.

To determine whether the wild type *ARC5* gene could complement the mutation, the predicted *ARC5* gene (a transgene containing the predicted At3g19730 /At3g19720 locus plus 1.9 kb and 1.1 kb of the 5' and 3' flanking DNA, respectively) was amplified from the DNA of BAC MMB12 by PCR using the primers 5'- GGAATTCCGAGTCGAGTTGCTTTGTTG-3' and 5'- CGTCTAGAGCTTACCTCAAAGGTACATGGA-3'. The PCR product was digested with *Eco*RI and ligated into a derivative of the transformation vector pLH7000 (http://www.dainet.de/baz/jb2000/jb_2000direkt.htm) digested with *Eco*RI and *Sma*I. The construct was transferred to *A. tumefaciens* GV3101 and introduced into *arc5* plants by floral dipping. The phenotypes of the T_1 plants were determined by microscopy. Microscopic analysis of T_1 transgenic plants indicated that the chloroplast division defect in the mutant was fully or partially rescued by the wild-type transgene.

Thus, from the results described above, which include the point mutation in At3g19730 /At3g19720 in *arc5*, complementation of the mutant phenotype by the wild-type gene, and ability of a fragment from At3g19730 /At3g19720 to confer an *arc5*-like phenotype in wild-type plants when expressed in the antisense orientation, indicate that the *ARC5* locus and At3g19730 /At3g19720 represent the same gene.

A cDNA for *ARC5* was isolated using RT-PCR. Based on the sequencing data and ORF analysis, primers were chosen to amplify a region from 93 bp upstream of the predicted start codon to 152 bp downstream of the stop codon. After the cDNA fragments were cloned into Bluescript KS+ vector, two distinct cDNAs encoding proteins with uninterrupted reading frames of 777 or 741 amino acids were found. These results indicate that the *ARC5* transcript is alternatively spliced. The longer cDNA contained a sequence that was spliced out of the shorter cDNA as the 15th intron; however, its presence in the longer cDNA did not interrupt the reading frame. Table 10 shows the SEQ ID NOs for ARC5 nucleic acids and proteins. The NCBI annotation is included in Table 10, as indicated.

The protein sequences were blasted against the NCBI protein database. The amino acid sequences of ARC5 were deduced from the cDNA sequence; the long form of the cDNA

encodes a protein of 777 amino acids and 87.2 kDa, whereas the shorter form of the cDNA encodes a protein of 741 amino acids and 83.5 kDa. The sequence alignment was performed with the CLUSTALW multiple alignment program (Thompson, J. D. *et al.* (1994) *Nucleic Acids Res.* 22, 4673-4680) at the Biology Workbench 3.2 website (http://biowb.sdsc.edu/). Protein sequences used for the phylogenetic analysis were aligned with Clustal X (Thompson, J. D. *et al.* (1997) *Nucleic Acids Res.* 25, 4876-4882) using default settings. Neighbor joining and maximum parsimony analyses were performed using PAUP version 4.0b10 (Swofford, D. L. (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10* (Sinauer Associates, Sunderland, Massachusetts)) with default settings except for ties being randomly broken. Neighbor-joining and maximum parsimony analyses produced topologically identical trees. Bootstrap analyses were performed on the neighbor-joining and maximum parsimony trees with one thousand replications. GENBANK® accession numbers for proteins aligned with ARC5 (longer form, accession no. AY212885) are as follows: human Dynamin-1 (NP_004399), yeast Dnm1p (NP_013100), At1g53140 (NP_175722), rice dynamin like protein (BAB56031), ADL6 (AAF22291), At5g42080 (NP_568602), *Glycine* phragmoplastin (AAB05992), tobacco phragmoplastin (CAB56619), At2g44590 (NP_181987), human Dynamin II (NP_004936), ADL2a (NP_567931), ADL2b (NP_565362), rice ADL2-like protein (BAB86118), worm Drp-1 (AAL56621) and human Dnm1p/Vps1p-like protein (JC5695).

The results, shown in Fig. 24, showed that the protein can be aligned over its entire length with numerous members of the dynamin family; most of the regions of the protein sequences can be aligned with the protein sequence of dynamin-I (GI# 4758182). Thus, the ARC5 protein contains three motifs found in other dynamin-like proteins: a conserved N-terminal GTPase domain, a pleckstrin homology (PH) domain shown in some proteins to mediate membrane association, and a C-terminal GTPase Effector Domain (GED) thought to interact directly with the GTPase domain and to mediate self-assembly (Danino, D. & Hinshaw, J. E. (2001) *Curr. Opin. Cell Biol.* 13, 454-460; and Hinshaw, J. E. (2000) *Annu. Rev. Cell Dev. Biol.* 16, 483-519). The shorter cDNA encoded a protein of 741 amino acids and 83.5 kDa identical to that of the larger gene product except for the absence of 36 amino acids encoded by the sequence of the 15th intron. These results suggest that the *ARC5*

transcript is alternatively spliced. Alternative splicing of dynamin genes in several other organisms has also been documented (Hinshaw, J. E. (2000) *Annu. Rev. Cell Dev. Biol.* 16, 483-519).

Phylogenetic analysis was performed to investigate the relationship between ARC5 and other members of the dynamin family of proteins. Only full-length sequences were used, though EST data indicate that related proteins are present in many plants and in green algae. ARC5 clustered with a group of proteins found in plants, but was in a distinct clade from other dynamin-like proteins in Arabidopsis with functions in cell-plate formation and mitochondrial division (Gu, X. & Verma, D. P. (1996) *EMBO J.* 15, 695-704; and Arimura, S.-i. & Tsutsumi, N. (2002) *Proc. Natl. Acad. Sci.* USA 99, 5727-5731). Surprisingly, the ARC5-like proteins clustered near ADL6, another Arabidopsis dynamin-like protein involved in vesicle trafficking from the trans-Golgi network to the vacuole in plants (37 Jin, J. B. *et al.* (2001) *Plant Cell* 13, 1511-1526).

Based on the similarity of ARC5 to dynamin and its relatives, ARC5 is contemplated to represent a new class of a dynamin-like proteins that functions specifically in chloroplast division.

The subcellular localization of ARC5 was investigated by expressing a GFP-ARC5 fusion protein in transgenic plants. The GFP sequence was amplified from plasmid smRS-GFP (Davis, S. J. & Vierstra, R. D. (1998) *Plant Mol. Biol.* 36, 521-528) with the primers 5'-CGGGATCCATGAGTAAAGGAGAAGAACT-3' and 5'-GCTCTAGATAGTTCATCCATGCCATGT-3'. The PCR product was digested with *Bam*HI and *Xba*I. The *ARC5* coding region and 1.1 kb of the 3' flanking DNA were amplified from the MMB12 BAC clone with primers 5'-GGACTAGTACGATGGCGGAAGTATCAGC-3' and 5'-CGGGATCCGCACCGAAGGAGCCTTTAGATT-3'. The PCR product was digested with *Spe*I and *Eco*RI. cDNA fragments encoding GFP and ARC5 were subcloned into Bluescript KS+ (Stratagene) that had been digested with *Eco*RI and *Bam*HI to create a *GFP-ARC5* fusion construct. The *ARC5* promoter was amplified from MMB12 with primers 5'-GACTAGTTGGCTCAACGCTTACCTCAA-3' and 5'-CGGGATCCGCCATCGTCTCTTACGA-3', and cloned into Bluescript KS+ (Stratagene) between the *Spe*I and *Bam*HI sites. The promoter fragment was then subcloned into the

plasmid containing the *GFP-ARC5* fusion construct at the 5' end of the fusion. The resulting plasmid was digested with *Spe*I and *Eco*RI, and the promoter-*GFP-ARC5* cassette was subcloned into a derivative of the transformation vector pLH7000 (http://www.dainet.de/baz/jb2000/jb_2000direkt.htm). The plasmid was transferred to *A. tumefaciens* GV3101 and used to transform wild-type *A. thaliana* plants (Col-0) as described above. The GFP-ARC5 localization pattern was visualized by fluorescence microscopy in T_1 plants. For *in vivo* detection of green fluorescent protein (GFP), fresh leaf tissue was mounted in water and viewed with an L5 filter set (excitation 455 nm to 495 nm, emission 512 to 575 nm) and a 100X oil immersion objective of a Leica DMR A2 microscope (Leica Microsystems, Wetzlar, Germany) equipped with epifluorescence illumination. Images were captured with a cooled CCD camera (Retiga 1350EX, Qimaging, Burnaby, British Columbia, Canada) and processed with Adobe Photoshop imaging software (Adobe Systems, San Jose, CA).

Because overexpression of chloroplast FtsZ proteins can result in a dominant-negative phenotype (Vitha, S. *et al.* (2001) *J. Cell Biol.* 153, 111-119), the native *ARC5* promoter was used to create the *GFP-ARC5* transgene for expression in wild-type plants (Col-0). Fluorescence microscopy showed that the fusion protein was localized in a ring-like pattern at the site of the chloroplast constriction. This ring could be faintly detected in unconstricted chloroplasts, suggesting that ARC5 may act at an earlier stage of division than previously hypothesized (Pyke, K. A. & Leech, R. M. (1994) Plant Physiol. 104, 201-207; and Robertson, E. J. *et al.* (1996) Plant Physiol. 112, 149-159). However, ARC5 is not required for FtsZ ring formation, the earliest known event in the assembly of the chloroplast division apparatus (Miyagishima, S. *et al.* (1999) Planta 207, 343-353; Miyagishima, S. *et al.* (2001) Plant Cell 13, 2257-2268; and 40 Bleazard, W. et al. (1999) Nature Cell Biol. 1, 298-304), since the FtsZ ring can be detected in the *arc5* mutant. The GFP-ARC5 fusion protein was most obvious in visibly constricted chloroplasts, perhaps as a consequence of ring thickening during constriction. Similar localization patterns have been described for FtsZ1 and FtsZ2 (Vitha, S. *et al.* (2001) J. Cell Biol. 153, 111-119).

Even though ARC5 mediates chloroplast division, it is not predicted by subcellular targeting prediction programs to be imported to the chloroplast. To further define the

topology of the ARC5-containing ring with respect to the chloroplast envelope membranes, *in vitro* chloroplast import and protease protection assays were employed. Transcription/translation reactions, chloroplast isolation, *in vitro* import reactions, proteolytic treatments, and post-import fractionation and analysis were performed as described (McAndrew, R. S. *et al.* (2001) Plant Physiol. 127, 1656-1666). The longer *ARC5* cDNA, after subcloning into Bluescript KS+ as described above, was used for these experiments.

A radiolabeled translation product corresponding to the longer *ARC5* cDNA was generated by coupled transcription/translation, then incubated with isolated pea chloroplasts. Subsequent fractionation of the chloroplasts indicated that the translation product was associated with the membrane fraction, but was not processed. The binding of the ARC5 translation product to isolated chloroplasts may be effected in part by the PH domain, which has been shown to mediate lipid binding of other dyanamin-like proteins (Hinshaw, J. E. (2000) Annu. Rev. Cell Dev. Biol. 16, 483-519; and 38 Lee, S. H. *et al.* (2002) *J. Biol. Chem.* 277, 31842-31849). In contrast, two chloroplast-targeted control proteins, one localized to the inner envelope and the other to the stroma, were processed upon import, consistent with the presence of N-terminal transit peptides, and associated with the membrane and soluble chloroplast fractions, respectively. In addition, the two control proteins were both protected from proteolysis by thermolysin, which does not penetrate the outer envelope (Cline, K. *et al.* (1984) Plant Physiol. 75, 675-678), whereas the ARC5 translation product was fully degraded by this protease. These data provide evidence that the ARC5-containing ring represented by the GFP-ARC5 fusion protein is situated on the cytosolic surface of the outer chloroplast envelope membrane. The position of ARC5 on the chloroplast surface is topologically equivalent to that of Dnm1p, a dynamin-like protein that mediates mitochondrial division in yeast (Bleazard, W. et al. (1999) Nature Cell Biol. 1, 298-304).

Blast searching indicates a second homologue of ARC5. It is predicted that this gene also functions in chloroplast division. This is based upon the observation of a slow but continued chloroplast division in *arc5*, which may be due to the presence of the second *ARC5* homologue (At1g53140) in a duplicated region of the *Arabidopsis* genome (Pyke, K. A. & Leech, R. M. (1994) *Plant Physiol.* 104, 201-207), and whose function might overlap that of

ARC5. Table 10 shows the coding and protein sequences for ARC5, as well as the NCBI and MIPS predicted protein sequence of the ARC5 homologue.

Table 10

ARC5

Gene	**SEQ ID NO**	**Figure Number**
ARC5 Genomic (BAC MMB12(GB:AP000417))	11	9
ARC5 cDNA	12	10
ARC5 Protein	13	11
NCBI ARC5 Genomic (BAC MMB12(GB:AP000417))	14	12
NCBI ARC5 cDNA	15	13
NCBI ARC5 Protein	16	14
NCBI ARC5 Homologue (protein)	17	15
MIPS ARC5 Homologue (protein)	18	16
ARC5 Genomic [1]	26; 27 [2]	24

Dynamin and its relatives are large GTPases that participate in a variety of organellar fission and fusion events in eukaryotes, including budding of endocytic and Golgi-derived vesicles, mitochondrial fission, mitochondrial fusion, and plant cell plate formation (reviewed in Danino, D. & Hinshaw, J. E. (2001) Curr. Opin. Cell Biol. 13, 454-460; and Hinshaw, J. E. (2000) Annu. Rev. Cell Dev. Biol. 16, 483-519). Dynamin has also been shown to regulate actin assembly and organization at membranes (Schafer, D. A. *et al.* (2002) Curr. Biol. 12, 1852-1857). ARC5 defines a new class of dynamin-like proteins that function specifically in plastid division, and its identification extends the range of cellular processes in which dynamin-like proteins participate.

EXAMPLE 7

Identification of Fzo-like plastid division gene

This Example describes the identification of an Fzo-like gene of Arabidopsis. A blast search of the Arabidopsis database using as the query sequence the yeast protein Fzo1, which functions in the control of mitochondrial morphology in yeast (Hermann et al 1998 J. Cell. Biol. 143:359; Rapaport et al. 1998 J. Biol. Chem. 273:20150; Sesaki and Jensen 1999 J. Cell. Biol. 147:699; Fritz et al. 2001 J. Cell Biol 152:683), revealed a related gene, designated Fzo-like gene, on chromosome 1, At1g03160 on BAC clone F10O3.

A Blast search of the Salk T-DNA insertion database identified 8 lines of Arabidopsis with T-DNA insertions in this gene. The seeds for these lines were obtained and germinated, and the resulting plants examined by microscopy for chloroplast division defects in leaves. Two lines exhibited abnormalities in chloroplast size and number, suggesting that At1g03160 functions in chloroplast division.

The open reading frame is predicted to contain a chloroplast transit peptide, further suggesting a role for in chloroplast division. Thus, Fzo-like protein is contemplated to possess several domains: a chloroplast transit peptide, a GTPase domain and two predicted trans-membrane domains. In Arabidopsis Fzo-like polypeptide, the predicted chloroplast transit peptide is the first 54 amino acids, the GTPase domain is between amino acids 350-500, and the two predicted trans-membrane domains are close to each other in the region between amino acids 770-830. EST information indicates that the 3' end of this gene probably resides in the neighboring BAC F15K9.

Knock-out of *AtFzo-like* results in impaired chloroplast development and division, and affects the growth and development of plant. Zero to ten chloroplasts of differing sizes are observed per cell in knock-out plants. The dumbbell-shape chloroplasts with constriction in the middle are frequently observed. The mutant plants looks yellow, smaller than wild type plants and flower later.

Localization experiments of AtFzo-like protein in the cell were performed as described above for ARC6, where AtFzo-like was fused to GFP. The results that AtFzo-like-GFP is localized to the vesicle-like structures associated with (or near) the chloroplast. The

level of AtFzo-like-GFP is positively correlated with the numbers of the vesicle-like structures.

Table 11 shows the SEQ ID NOs for the Fzo-like nucleic acid and protein sequences. Both the MIPS and the NCBI cDNA and translations are provided.

Table 11

Fzo-Like Gene

Gene	**SEQ ID NO**	**Figure Number**
MIPS Fzo Genomic	19	17
MIPS Fzo cDNA	20	18
MIPS Fzo Protein	21	19
NCBI Fzo Genomic	22	20
NCBI Fzo cDNA	23	21
NCBI Fzo Protein	24	22
3' Fzo Genomic (BAC F15K9)	25	23

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in chemistry, and molecular biology or related fields are intended to be within the scope of the following claims.